

X-Ray Crystallographic Studies on
Electron Transfer Proteins;
Rubredoxin from *Pyrococcus furiosus*,
Nitrogenase MoFe from *Azotobacter vinelandii* and
Ru(2,2'-bppy)₂(imd)His83 Azurin from *Pseudomonas*
aeruginosa

Thesis by
Michael W. Day

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

1996
(Submitted June 19, 1995)

Acknowledgments

As is the case with any scientific investigation, the projects described herein are the result of input from numerous sources. I would like to thank Professor Douglas C. Rees for taking me into his research group and for his continuous support, as both a scientific and personal mentor. Without his attentive direction during these studies and his constant fostering of a productive environment in which to conduct this research, these projects would have suffered greatly.

I would like to thank Dr. Leemor Joshua-Tor for her assistance, particularly during the low temperature data collection on rubredoxin, Barbara Hsu for her constant vigilance with the equipment and I thank Dr. Arthur Chirino for his assistance with the programs and techniques unique to macromolecular crystallography.

I extend my appreciation to all of the members of the Rees lab for their constant contribution and team spirit. In particular I thank Tim McPhillips for his insightful scientific and personal discussions (the later of which helped me maintain my sanity), Michael Stowell for his ever-ready willingness to be of assistance and to Debbie Woo for her help with the computers and many of the logistics involved in navigating this research.

Of course I cannot adequately thank the members of my family. Without their unfailing support and their constant presence in times of both clear sailing and times of crisis, this adventure would have never reached this point. I offer a special thanks to my fiancée, Elisabeth Stewart, for her invaluable and loving support during the writing of this thesis.

This thesis is dedicated to those people who contributed so much to my efforts while knowing all along that they would receive none of the glory.

THANK YOU!!!!!!!!!!!!

"... from Him and through Him and to Him are all things.
To Him be the glory forever! ..."

Abstract

The x-ray crystal structure of the oxidized and the reduced forms rubredoxin from *Pyrococcus furiosus*, a hyperthermophilic marine *Archae*, have been solved by molecular replacement and refined by the method of restrained least squares to a maximum resolution of 1.1 Å for the oxidized form and 1.5 Å for the reduced form. The oxidized form of the protein crystallizes in the orthorhombic space group $P 2_1 2_1 2_1$ with unit cell dimensions of $a=33.8\text{Å}$, $b=34.6\text{Å}$, $c=43.4\text{Å}$ and $V=50,755\text{Å}^3$. The reduced form crystallizes in the same space group with the nearly identical unit cell dimensions of $a=33.8\text{Å}$, $b=34.5\text{Å}$, $c=43.2\text{Å}$ and $V=50,375\text{Å}^3$. Data on both forms was collected at -161°C . Three refinement packages were used in the refinement and the results from each are discussed as are the possible determinants of the thermal stability. Refinement of the oxidized form (414 protein atoms and 104 solvent oxygens) with TNT or XPLOR resulted in a crystallographic residual of approximately 17% and a model with rms deviations of bond distances and angles from target values of approximately 0.015 Å and 2.5° respectively. Refinement of the oxidized form with SHELXL-93 resulted in a model with 132 solvent oxygens and an $R=13.9\%$ ($R_{\text{free}}=17.2\%$) and $\text{GOF}=1.08$. The rms deviation from the target values for bond distance and angles are 0.014 Å and 1.75° respectively. Refinement of the reduced form with TNT (110 solvent oxygens) or SHELXL-93 (including 130 solvent oxygen atoms) results in an R-factor of approximately 17% and the geometry of the model deviates from the target values by approximate rms values of 0.022 Å for the bond distances and 3.0° for the bond angles.

The x-ray crystal structure of the MoFe nitrogenase protein from *Azotobacter vinelandii* has been refined against data collected at the Stanford Synchrotron Radiation Laboratory (SSRL). The data extends to a maximum resolution of 2.2 Å and two packages were used in the restrained least squares refinement. Refinement of the model (including 625 solvent oxygens) with TNT or XPLOR yielded a crystallographic residual of less than 18% and a model with bond distances and angles deviate from the target values by rms values of 0.02 Å and 2.5° respectively.

The x-ray crystal structure of the $\text{Ru}(2,2'\text{-bipyridine})_2(\text{imidazole})\text{His83}$ azurin from *Pseudomonas aeruginosa* has been solved and refined by the method of restrained least squares to a limiting resolution of 2.5 Å. The labeled protein crystallizes in the monoclinic space group $C 2$ with $a=100.6\text{Å}$, $b=35.4\text{Å}$, $c=74.7\text{Å}$, $\beta=106.5^\circ$, $V=255,069\text{Å}^3$ and $Z=8$. Data was collected at -161°C to a maximum resolution of 2.3 Å yielding a data set that is 82% complete containing 11,083 reflections. Refinement in TNT (including 150 solvent oxygen atoms) resulted in an R-factor of 17.3% with rms deviations in the model bond distances and angles from ideal values of 0.026 Å and 3.09° respectively.

Table of Contents

Abstract	iii
List of Figures	v
List of Tables	xii
Introduction to Electron Transfer Proteins	1
Chapter 1 Rubredoxin from <i>Pyrococcus furiosus</i>	9
1.1 Introduction to Rubredoxin	10
1.2 Experimental	19
1.2.1 Crystalization	19
1.2.2 Data Collection	21
1.2.3 Structure Determination	24
1.2.4 Low Resolution Refinement	28
1.2.5 High Resolution Refinement	36
1.2.6 High Resolution Refinement with SHELXL-93	44
1.3 Structure Discussion	59
1.4 Possible Determinants of Thermostability	82
Chapter 2 Nitrogenase Mo-Fe Protein from <i>Azotobacter vinelandii</i>	93
2.1 Introduction to Nitrogenase	94
2.2 Experimental	103
2.2.1 Structure Solution	103
2.2.2 Structure Refinement	106
2.2.3 M-Cluster Composition	119
2.2.4 P-Cluster Composition	125
2.3 Structure Discussion	138
2.4 Implications for Nitrogen Reduction	161
Chapter 3 Azurin from <i>Pseudomonas aeruginosa</i>	170
3.1 Introduction to Azurin	171
3.2 Experimental	178
3.3 Structure Discussion	185
Appendix 1 TNT Nitrogenase MoFe cofactor Geometry Definition	194
Appendix 2 XPLOr Nitrogenase MoFe cofactor Geometry Definition	199
Appendix 3 TNT Ru(2,2'-bppy) ₂ (imd)His83 Geometry Definition	208

List of Figures

		Page
Figure 1.1-1	A schematic representation of the metabolic pathways in <i>P. furiosus</i> .	11
Figure 1.1-2	Labeling patterns for glycolysis showing the eventual fate of carbon.	13
Figure 1.1-3	Sequence alignment for the rubredoxins from anaerobic bacterium.	16
Figure 1.2.3-1	The Harker sections of the patterson map calculated from the high resolution data using all data between 2.5Å and 1.1Å.	25
Figure 1.2.3-2	The Harker sections of the patterson map calculated from the anomalous differences for all data between 4-1.0Å.	26
Figure 1.2.3-3	F observed map calculated from phases derived from the iron position.	27
Figure 1.2.4-1	Average main chain and side chain temperature factors for the oxidized form.	29
Figure 1.2.4-2	Stereo view of the final 2Fo – Fc electron-density map in the region around Glu 14 of the oxidized form.	29
Figure 1.2.4-3	Dependence of R-factor on resolution for the oxidized form of rubredoxin.	30
Figure 1.2.4-4	Ramachandran plot for the oxidized form of RdPf.	31
Figure 1.2.4-5	Stereo view of the 2Fo – Fc electron density map for the reduced form of RdPf in the area near Trp-3, Glu-14 and Phe-29.	33
Figure 1.2.4-6	Average main chain and side chain temperature factors for the oxidized form.	33
Figure 1.2.4-7	Ramachandran plot of the reduced form of RdPf.	34
Figure 1.2.4-8	Dependence of R-factor on resolution for the reduced form of rubredoxin.	34
Figure 1.2.5-1	Stereo view at the 5σ level of 2Fo-Fc electron density for the Fe-S center in rubredoxin showing the anisotropy of the density.	37
Figure 1.2.5-2	Average main chain and side chain temperature factors for the model of oxidized rubredoxin after refinement in XPLOR at between 10-1.1Å.	37
Figure 1.2.5-3	Stereo view of refined 2Fo-Fc the electron density map after refinement in XPLOR at 1.1Å resolution.	38
Figure 1.2.5-4	R-factor as a function of resolution after refinement in XPLOR.	39

Figure 1.2.5-5	Average main chain and side chain temperature factors for the model of oxidized rubredoxin after refinement in TNT between 10-1.1 Å.	39
Figure 1.2.5-6	Plot of R-factor as a function of resolution from refinement with TNT.	41
Figure 1.2.5-7	Stereo view of representative electron density in the 2Fo-Fc map from the final model in the area around Trp-3, Glu-14 and Phe-29.	42
Figure 1.2.5-8	Average temperature factors for the model of reduced rubredoxin refined in TNT.	42
Figure 1.2.5-9	Plot of R-factor as a function of resolution from refinement with TNT.	43
Figure 1.2.6-1	Stereo view of the rubredoxin after unrestrained refinement in SHELXL-93 showing that to a large degree the molecule is intact and visually adequate.	46
Figure 1.2.6-2	Stereo view of the electron density map around Tyr-12 after refinement in SHELXL-93.	51
Figure 1.2.6-3	Plot of average equivalent B-factors for the oxidized form of RdPf after refinement in SHELXL-93 with anisotropic displacement parameters and hydrogen atoms included.	52
Figure 1.2.6-4	a) Electron density from the 2Fo-Fc map around Trp-3, Glu-14, Phe-29 and Lys-50 at the 2σ level. b) ORTEP representation of the same region showing the thermal ellipsoids at the 60% level.	53
Figure 1.2.6-5	a) Electron density at 4σ from the 2Fo-Fc map for residues in the hydrophobic core. b) ORTEP drawing of the hydrophobic core showing 60% atomic ellipsoids.	54
Figure 1.2.6-6	A plot of R-factor versus resolution for the SHELXL-93 refined model of RdPf.	55
Figure 1.2.6-7	Plot of average temperature factors for the reduced form of RdPf after refinement in SHELXL-93 at 8-1.5 Å resolution.	56
Figure 1.2.6-8	Estimation of average coordinate error from a plot of R-factor versus resolution for the reduced form of rubredoxin after refinement between 8-1.5 Å with SHELXL-93.	57
Figure 1.2.6-9	Stereo view of the electron density at 2s from the 2Fo-Fc map in the area around the residues believed to play an important role in the thermal stability of RdPf, Trp-3 and Glu-14.	57
Figure 1.3-1	Molscript representation showing the overall topology of the rubredoxin from <i>Pyrococcus furiosus</i> .	59
Figure 1.3-2	Stereo view of the Cα trace showing the overall folding of RdPf.	60

Figure 1.3-3	Stereo view of the hydrogen-bonding network in the β -sheet including the hydrogen bond, O5...N9, that defines a G-1 β -bulge.	60
Figure 1.3-4	Stereo view of the aromatic residues and Leu 32 that form the hydrophobic core of RdPf.	61
Figure 1.3-5	Plot of the bond distances between iron and sulfur showing the respective esd's.	65
Figure 1.3-6	Stereo view of the pseudo-twofold axis around the iron-sulfur cluster. NH...S bonds are shown.	65
Figure 1.3-7	Plot of the root mean square differences in the backbone atoms of the various rubredoxin crystal structures after superposition on the deposited coordinates of <i>P. furiosus</i> (1CAA).	68
Figure 1.3-8	Side chain to main chain hydrogen bond from Lys-45, Phe-29 and Leu-32 that occurs in all known rubredoxins.	68
Figure 1.3-9	Molscript drawing of the residues unique to the rubredoxin from <i>Pyrococcus furiosus</i> .	71
Figure 1.3-10	Stereo view of the hydrogen bonds and salt bridge formed by Glu 14.	72
Figure 1.3-11	Sequences of RdPf and the four other rubredoxins whose x-ray crystal structure is known. The cystien residues are shown in bold face, \uparrow shows the location of the conserved solvent molecules, \downarrow denotes residues unique to RdPf and marks residues that are strictly conserved in all known rubredoxin sequences. The symbols along the bottom denote elements of secondary structure; \blacktriangle represents strands of the β -sheet and $\#$ represents the helical corners.	76
Figure 2.2.2-1.	Plot of average temperature factors vs. residue number for MoFe after refinement with TNT.	109
Figure 2.2.2-2.	Plot of the R-factor from TNT refinement as a function of the resolution.	110
Figure 2.2.2-3.	Stereo view of representative electron density of the Mofe protein after refinement with TNT.	110
Figure 2.2.2-4.	Ramachandran plot of the ϕ , ψ angles for each subunit of MoFe.	111
Figure 2.2.2-5.	Plot of average B-factors vs. residue number for each of the four subunits of Av1 after refinement with XPLOR.	112
Figure 2.2.2-6.	A Luzzati plot of the R-factor from the XPLOR refinement vs. the resolution.	113
Figure 2.2.2-7.	Stereo view of $2F_o - F_c$ electron density from the same helix shown in Figure 2.2.2-3. This density is representative of the entire protein after refinement with XPLOR. The map is contoured at the 2σ level.	113

Figure 2.2.2-8.	Ramachandran plot of the ϕ , ψ torsion angles for each subunit of Av1 after refinement with XPLOR.	114
Figure 2.2.3-1.	The current model of the MoFe cofactor without any of the ligands to the polypeptide chain or homocitrate.	119
Figure 2.2.3-2.	The model of the M-cluster proposed by Bolin et al. with a sulfur atom at the octahedral site in the middle of the cluster.	122
Figure 2.2.3-3.	Stereo view of the electron density from the 2Fo-Fc map from the current model of the M-cluster after refinement with TNT.	123
Figure 2.2.4-1.	Stereo view of the current model for the P-cluster pair.	125
Figure 2.2.4-2.	Stereo view of the P-cluster pair showing all peaks in the Fo-Fc map above the 12σ level and superimposed on the coordinates of the deleted atoms.	127
Figure 2.2.4-3.	Stereo view of the 8σ electron density of the 2Fo-Fc map calculated from the same model used in the map shown in Figure 2.2.4-1 above. The P-cluster pair and ligand coordinates are superimposed on map for reference.	127
Figure 2.2.4-4.	Stereo view of the peaks above 12σ in the Fo-Fc map calculated using the model with the P-cluster or its ligands after refinement with TNT.	128
Figure 2.2.4-5.	Stereo view of the 2Fo-Fc electron density map from the same coordinates as used above. Contours correspond to 8σ electron density.	128
Figure 2.2.4-6.	Stereo view of the Fo-Fc electron density map after refinement of the model with the iron atoms of the P-cluster pair. The density is shown at the 6σ level.	130
Figure 2.2.4-7.	Stereo view of the electron density at 5σ of the 2Fo-Fc map from the refinement above. The map is superimposed on the P-cluster and its ligands.	130
Figure 2.2.4-8.	Stereo view of the Fo-Fc map at 6σ and 2.8\AA . Significant density can be seen between the positions that would be occupied by S(1A) and S(1B).	131
Figure 2.2.4-9.	Stereo view of the Fo-Fc electron density at 6σ after refinement with the 2.7\AA data collected in house. Significant electron density can be seen along the S(1A) and S(1B) bond vector.	132
Figure 2.2.4-10.	Stereo view of the electron density from the 2Fo-Fc map at 2σ . This map was calculated after refinement against the low resolution data set. Notice the absence of electron density between S(1A) and S(1B).	132
Figure 2.2.4-11.	Stereo view of 5σ electron density from the Fo-Fc map after S(1A) is deleted and the coordinates are refined.	133

Figure 2.2.4-12.	Stereo view of electron density from the 2Fo-Fc map around the P-cluster pair after deletion of S(1A) and refinement of the model. The map is contoured at the 4σ level.	133
Figure 2.2.4-13.	Stereo view of 5σ electron density from the Fo-Fc map after S(1B) is deleted and the coordinates are refined.	134
Figure 2.2.4-14.	Stereo view of electron density from the 2Fo-Fc map around the P-cluster pair after deletion of S(1A) and refinement of the model. The map is contoured at the 4σ level.	135
Figure 2.2.4-15.	Stereo view of 2Fo-Fc electron density map after averaging of electron density with the entire P-cluster pair and all liganding atoms deleted from the model. The map is contoured at 5σ .	136
Figure 2.2.4-16.	Stereo view of 2Fo-Fc electron density map after averaging of electron density with the sulfur atoms of the P-cluster pair deleted from the model. The map is contoured at 3σ .	137
Figure 2.3-1.	Molscript drawing of the α -subunit of Avl showing the three domains, I, II and III.	140
Figure 2.3-2.	A schematic diagram of the secondary structure of the MoFe α -subunit. The three domains are labeled I, II and III. The positions of ligands to the M-cluster and the P-cluster are designated by M and P respectively.	141
Figure 2.3-3.	Molscript drawing of the β -subunit of MoFe.	142
Figure 2.3-4.	A schematic representation of the β -subunit of MoFe. The three domains are labeled I', II' and III'.	143
Figure 2.3-5.	Molscript drawing of the $\alpha\beta$ subunit pair. The view is down the approximate two-fold axis, passing through the P-cluster pair, that relates the α and β subunits.	132
Figure 2.3-6.	Molscript drawing of the complete tetramer. The view is down the z-axis along the approximate non-crystallographic two-fold that relates the two $\alpha\beta$ -subunits.	144
Figure 2.3-7.	The cation binding site at the interface of the two $\alpha\beta$ -subunit pairs.	147
Figure 2.3-8.	The MoFe-cofactor and its ligands. The hydrogen bond between the O γ of Ser α 278 and the S γ of Cys α 275 is shown by the dashed line.	148
Figure 2.3-9.	The M-cluster showing atom labeling scheme.	150
Figure 2.3-10.	Stereo view of the P-cluster and its ligands.	152
Figure 2.3-11.	The P-cluster pair showing the atom labeling scheme.	153

Figure 2.4-1.	Stereo view of a space filling model of a cleft in the protein surface (made up of residues $\alpha 45$ - $\alpha 52$, $\alpha 185$ - $\alpha 203$, $\alpha 274$ - $\alpha 285$, $\alpha 354$ - $\alpha 362$ and $\alpha 377$ - $\alpha 389$) that funnels into a solvent filled channel (lined by residues Ser α -192, His α -195, His α -196, Arg α -277, Tyr α -281 and His α -383) leading to the FeMo-cofactor.	161
Figure 2.4-2.	Stereo view of a space filling model of the cleft in the surface of the α -subunit formed by $\alpha 260$ - $\alpha 288$, $\alpha 293$ - $\alpha 315$, $\alpha 350$ - $\alpha 368$ and $\alpha 378$ - $\alpha 394$.	162
Figure 2.4-3.	Nitrogen bound to the face of the FeMo-cofactor at a four iron face with two bridging sulfur atoms folded down and away making the face accessible.	166
Figure 2.4-4.	Dinitrogen binding within the central cavity of the FeMo-cofactor.	167
Figure 3.1-1.	Stereo view of the Cu site in the azurin from <i>Pseudomonas aeruginosa</i> . The three amino acids His 46, Cys 112 and His 117 contribute equatorial ligands, and the amino acids Gly-45 and Met 121 supply the axial ligands.	171
Figure 3.1-2.	Stereo view of the $\text{Ru}(2,2'\text{-bppy})_2(\text{imd})_2$ used to label azurin at His83. His83 displaces one imadazole to form a covalent complex.	174
Figure 3.2-1.	Plot of average temperature factors vs. residue number.	181
Figure 3.2-2.	Stereo view of representative electron density for the azurin from <i>P. aeruginosa</i> .	182
Figure 3.2-3.	Ramachadran plot of the (Φ , Ψ) angles for labeled azurin. Residues represented by + are glycines.	182
Figure 3.2-4.	Plot of the dependence of R-factor on resolution for $\text{Ru}(2,2'\text{-bppy})_2(\text{imd})$ labeled azurin. Theoretical curves are shown for an estimated error of 0.25Å (top) and 0.20Å (bottom). Only reflections between 7 – 2.5Å resolution were used in the refinement.	183
Figure 3.3-1.	Moscript representation of the azurin from <i>P. Aeruginosa</i> showing the overall folding topology of a crystallographically unique monomer.	185
Figure 3.3-2.	Stereo view of the alpha carbon trace illustrating the overall fold. This is the standard view with the copper atom at the northern end with the ligands to copper labeled. The N-terminal and C-terminal residues are labeled at the bottom of the molecule.	186
Figure 3.3-3.	Molscript drawing of the dimer of crystallization. The copper atom in the active site can be seen inside and at one end of each β -barrel. The $\text{Ru}(2,2'\text{-bppy})_2(\text{imd})$ complex can be seen at the interface of the two molecules. The copper of crystallization is visible near the N-terminus of each molecule.	187

Figure 3.3-4.	Stereo view of a superposition of the copper site in the unlabeled azurin on the ruthenium labeled azurin. The thick lines represent the ruthenium labeled azurin and the thin lines represent the structure described by Nar et al.	188
Figure 3.3-5.	Stereo view of an alpha carbon trace of the azurin molecule (in the standard orientation) showing the placement of the Ru(2,2'-bppy) ₂ (imd)His83 moiety.	189
Figure 3.3-6.	Stereo view of Ru(2,2'-bppy) ₂ (imd)His83. Residues 82 thru 85 are included. Note the hydrogen bond formed between the carbonyl oxygen of Thr84 and the Nε2 nitrogen of the imidazole ring in His83.	190
Figure 3.3-7.	Stereo view of the electron density from the 2Fo-Fc map around His83.	190
Figure 3.3-8.	Stereo view of a superposition of the ten amino acid stretch of residues 80-90 from wild type azurin, an the ruthenium labeled azurin. The thick lines represent the labeled azurin and the thin lines represent the wild type azurin solved by Nar et al.	191
Figure 3.3-9.	Stereo view of the copper near the N-terminus.	192

List of Tables

		Page
Table 1.2.2-1	Data collection statistics for the four data sets used in the refinement of RdPf.	21
Table 1.2.2-2	Data was collected on crystals of approximately 0.3 mm ³ in size using the following orientations, scanning in Ω at the 2θ and Φ angles shown with a crystal to detector distance of 8.0 cm.	22
Table 1.2.6-1	Results of the evaluation of refinement methods in SHELXL-93.	48
Table 1.3-1	Bond distances (Å) and bond angles (degrees) in the iron-sulfur cluster after refinement of the coordinates with XPLOR or TNT.	63
Table 1.3-2	Bond distances (Å) for the iron-sulfur center for all rubredoxin coordinate sets in the Brookhaven data bank (top) and for the current refinement of RdPf with SHELXL-93 (bottom).	64
Table 1.3-3	Hydrogen bond distances (Å) for sulfur to nitrogen in the coordinates refined with SHELXL-93.	70
Table 1.3-4	Solvent atom protein contacts in RdPf that occur in at least one other rubredoxin crystal structure deposited in the PDB.	77
Table 1.3-5	Solvent oxygens within H-bonding distance of the oxidized form of RdPf.	78
Table 1.3-6	Solvent oxygens within H-bonding distance of the reduced form of RdPf.	79
Table 1.4-1	Buried and exposed accessible surface areas for rubredoxins.	87
Table 2.2.2-1	Summary of refinement progress showing the results from each cycle for each refinement package used as well as the total number of atoms and number of solvent oxygens in the model.	107
Table 2.2.2-2	The overall average temperature factors (Å ²) for main chain and side chain atoms of each subunit of MoFe resulting from refinement in TNT or XPLOR.	109
Table 2.2.2-3	Bonds distances within the M-cluster and to its ligands. The initial lengths in the right-hand column are for 1MIN.	115
Table 2.2.2-4	Selected distances within the molybdenum-iron cofactor.	116
Table 2.2.2-5	P-Cluster bond distances.	117
Table 2.2.2-6	Selected P-Cluster distances.	117
Table 2.2.3-1	The results of the investigation into the identity of the Y ligand.	120

Table 2.2.4-1	The atomic positions for the Fe atoms of the P-cluster pairs in the pentultimate model are listed in the first column for comparison.	126
Table 2.2.4-2	The original iron positions are listed in the left-hand column and the positions of the peaks from the difference maps after refinement of the model against the combined data set are shown in the right-hand column.	129
Table 2.3-1	Table of selected distances within the MoFe-cofactor. The average distances result from averaging distances in both subunits from the results obtained by refinement in TNT and in XPLOR.	148
Table 2.3-2	Bond distances within the MoFe-cofactor and to its ligands. The averaged distances are the result of averaging the distances in both subunits from both refinement programs.	149
Table 2.3-3	Hydrogen bonds to the MoFe-cofactor and its ligands.	152
Table 2.3-4	Selected interatomic distances within the P-cluster pair. Initial distances are the average over both subunits of IMIN.	153
Table 2.3-5	Table of bond distances within the P-cluster pair and to its ligands. The average distances are calculated both the [A] and [D] subunits within and between TNT and XPLOR refinement results.	154
Table 2.3-6.	A list of the amino acid residues that line the polypeptide pocket of the α -subunit where the FeMo-cofactor is bound.	156
Table 2.3-7.	Table of the residues lining the cavity in the β -subunit that is analogous to the FeMo-cofactor binding pocket of the α -subunit.	157
Table 2.3-8.	The subunit interface regions that contain significant amounts of solvent.	158
Table 3.3-1.	Copper ligand distances (\AA) for the high resolution azurin crystal structures reported to date.	188
Table 3.3-2.	Table of bond distances (\AA) to the ruthenium atom of $\text{Ru}(2,2'\text{-bppy})_2(\text{imd})\text{His83}$ showing the values for both molecule A and B with the crystallographically determined distances for $\text{Ru}(2,2'\text{-bppy})_2(\text{imd})_2$.	191
Table 3.3-3.	Table of distances to the two copper ions found at the N-terminus.	192

**Introduction
to
Electron Transfer Proteins**

The three dimensional structure of electron transfer proteins is an active area of research. The thermodynamic and kinetic properties of the biological electron transfer process have been developed by Marcus et al. (Marcus and Sutin, 1985) and a balanced review of the models for electron transfer in biological systems has been published (Canters and van de Kamp, 1992). The interested reader is referred to these two papers since the electron transfer theory is not be the focus of this thesis. There are three types of redox cofactors found in proteins; small organic molecules, amino acid side chains or metal redox centers. The disulfide bridge of cystine is an example of amino acid side chains that can function as a redox center. Examples of small organic molecules that can mediate electron transfer in biological systems are quinones, nictotinamide adenine dinucleotide phosphate (NADP) and nictotinamide adenine dinucleotide (NAD). There are four basic types of electron transfer proteins; cytochromes, iron-sulfur proteins, blue copper proteins and flavodoxins (Adman, 1979; Meyer and Cusanovich, 1989). This thesis will focus on the structure of three metalloproteins, two iron-sulfur proteins (Rubredoxin and Nitrogenase MoFe) and one blue copper enzyme (Azurin).

In a review of the available crystal structures of electron transfer proteins Adam pointed to six properties that were common to electron transfer proteins (Adman, 1979). They are:

1. possession a cofactor that acts as an electron sink;
2. placement of this cofactor close enough to the surface of the protein to allow the entry or exit of an electron;
3. substantial alteration of the reduction potential of the cofactor by the protein;
4. the existence of a hydrophobic shell adjacent to but not necessarily surrounding the cofactor;
5. only small structural changes in the protein upon the transfer of

electrons;

6. a flexible architecture that permits expansion or contraction in preferred directions upon electron transfer.

Some metalloproteins simply transfer electrons (i.e., rubredoxin and azurin) while others may transport and/or activate small molecule (i.e., hemoglobin) and still others may both transfer electrons and activate small molecules (i.e., the nitrogenase MoFe protein). The transfer of electrons and activation of a small molecule generally requires multiple redox centers. Other examples of metalloenzymes that contain multiple redox sites and facilitate intramolecular electron transfer between metal centers that may have ligands bound to them are cytochrome *c* oxidase and xanthine oxidase.

The most simple of all electron transfer proteins are the rubredoxins. Rubredoxin is primarily found in anaerobic bacteria but their function is still unknown. These proteins are small (approximately fifty amino acids, molecular weight 6 kD) and contain a single iron atom liganded to four cysteine sulfurs in an approximately tetrahedral environment. The reduction potential of the Fe(III/II) couple is significantly lower in the protein than in water. For example, the reduction potential of the rubredoxin from *Clostridium pasteurianum* is -0.057 mV at pH 7 (Lovenberg and Sobel, 1965). Presumably the difference between that value and 0.77 V for $\text{Fe}^{+3}/\text{Fe}^{+2}$ in water arises from a stabilization of the Fe^{+3} form by the negative charge on the cysteinyl sulfurs (Eaton and Lovenberg, 1973). Comparative studies on these most simple of the redox proteins hold the possibility of understanding how the protein environment influences the electron transfer and redox properties of the iron site. For instance, it has been proposed that the conserved aromatic residues of the hydrophobic core of rubredoxin play an important role in mediating the transfer of electrons (Frey, Sieker et al., 1987; Adman, Sieker et al., 1991; Sun, Ueyama et al., 1993). Numerous studies on model compounds of the rubredoxin FeS_4 center have been reported (Werth, Kurtz et al., 1989; Sun, Ueyama et al., 1991; Walters, Dewan et al., 1991; Maelia,

Millar et al., 1992; Ueyama, Sun et al., 1992; Huang, Moura et al., 1993; Sun, Ueyama et al., 1993). A more complete description of the rubredoxin from the hyperthermophilic marine organism *Pyrococcus furiosus* is given in Chapter 1.

Ferredoxins are another class of iron-sulfur proteins. The simplest of the ferredoxins have a molecular weight of between 10-20 kD, are generally found in plant chloroplasts and contain a [2Fe-2S] redox center. The redox potential of these proteins is generally low at approximately -400 mV. Small ferredoxins (MW 6-10 kD) with four-iron clusters are found in many species of bacteria. These redox centers contain four iron atoms and four sulfides that occupy the alternating corners of a distorted cube. The [4Fe-4S] clusters of these proteins also have reduction potentials of approximately -400 mV. The modeling of [4Fe-4S] clusters and investigation of their electron transfer properties has attracted much attention (Excoffon, Laugier et al., 1991; Kodaka, Tomohiro et al., 1991; Ohno, Ueyama et al., 1991; Langen, Jensen et al., 1992; Roth and Jordanov, 1992; Ueyama, Oku et al., 1992; Zhou, Scott et al., 1992; Kambayashi, Nagao et al., 1993; Yanada, Nagano et al., 1993; Evans and Newton, 1994). The nitrogenase MoFe protein contains two different types of metal centers that are variations of the Fe₄S₄ cubane motif. As with the FeS₄ and Fe₄S₄ redox centers, the metal centers of the nitrogenase MoFe protein were an area of intense interest for investigators who were attempting to synthesize models of these metal centers (Coucouvannis, Challen et al., 1989; Challen, Koo et al., 1990; Challen, Koo et al., 1990; Cen, MacDonnell et al., 1994; Demadis and Coucouvannis, 1995; Laughlin and Coucouvannis, 1995; Malinak, Demadis et al., 1995). However, unlike the FeS₄ and Fe₄S₄ clusters, the x-ray crystallographic structure of these sites had not been solved until the report by Kim and Rees (Kim and Rees, 1992) and to date there have been no reports of the successful synthesis of these redox centers. The 2.2Å resolution refinement of the crystallographically determined structure of the nitrogenase MoFe protein is reported in Chapter 2.

Blue copper proteins also serve as electron-transfer agents in biological systems. These proteins contain three ligands (two His and one Cys) are tightly bound to the copper atom. Two other ligands (Met and a mainchain carbonyl oxygen) occupy the axial positions in a trigonal bipyramidal coordination sphere around the copper atom. The intense blue color of these proteins arises from the Cu-S interaction and the reduction potential is believed to be tuned by the Met sidechain. The reduction potential of the blue copper proteins tend to be approximately 300 mV, a value much higher than that observed in the iron-sulfur proteins. Bioinorganic chemists have used redox-active inorganic complexes as artificial substrates in an attempt to understand the electron transfer process through the blue copper proteins. Chapter 3 will discuss the structure of azurin with $\text{Ru}(2,2'\text{-bppy})_2(\text{imd})$ as a ligand to His83. This structure has relevance to the work currently in progress in the laboratory of Professor H. B. Gray here at Caltech (Gray and Solomon, 1981; Gray and Malmström, 1983; Kostic, Margalit et al., 1983; Gray, 1986; Che, Margalit et al., 1987; Mizoguchi, Di Bilio et al., 1992; Winkler and Gray, 1992).

References

- Adman, E. T. (1979). "A Comparison of the Structures of Electron Transfer Proteins." *Biochimica et Biophysica Acta*, **549**, 107-144.
- Adman, E. T., Sieker, L. C. and Jensen, L. H. (1991). "Structure of Rubredoxin from *Desulfovibrio vulgaris* at 1.5 Å Resolution." *Journal of Molecular Biology* **217**, 337-352.
- Canthers, C. W. and van de Kamp, M. (1992). "Protein Mediated Electron Transfer." *Curr. Opin. Struc. Biol.* **2**, 859-869.
- Cen, W., MacDonnell, F. M., Scott, M. J. and Holm, R. H. (1994). "Heterometal Clusters Containing the Cuboidal Fe_4S_3 Fragment: Synthesis, Electron Distribution, and Reactions." *Inorg. Chem.* **33**, 5809-5818.
- Challen, P. R., Koo, S.-M., Dunham, W. R. and Coucouvanis, D. (1990). "New $\mu_2\text{-S}^{2-}$ -Coupled, Singly Bridged Double Cubane with the $[\text{Fe}_4\text{S}_4\text{Cl}_3]_2\text{S}]^{4-}$ Core The Stepwise Synthesis and Structural Characterization of $(n\text{-Bu}_4\text{N})_2(\text{Ph}_4\text{P})_2[\text{Fe}_4\text{S}_4\text{Cl}_3]_2\text{S}]$." *J. Am. Chem. Soc.* **112**, 2455-2456.

- Challen, P. R., Koo, S.-M., Kim, C. G., Dunham, W. R. and Coucouvanis, D. (1990). "Nitrogenase Substrates as Intercluster Bridging Units between the Mo Atoms in Doubly Bridged, Double Cubanes. The Synthesis and Characterization of the $[\text{MoFe}_3\text{S}_4\text{Cl}_2(\text{Clcat})]_2(\mu_2\text{-S})(\mu_2\text{-L})]^{n-}$ Anions ($\text{L} = \text{N}_2\text{H}_4$, $n=4$; $\text{L} = \text{CN}^-$, $n=5$)." *J. Am. Chem. Soc.* **112**, 8606-8607.
- Che, C. M., Margalit, R., Chiang, H. G. and Gray, H. B. (1987). "Ruthenium Modified Proteins - Reactions of $\text{cis-}[\text{Ru}(\text{NH}_3)_4(\text{H}_2\text{O})_2]^{+2}$ and $\text{cis-}[\text{Ru}(\text{en})_2(\text{H}_2\text{O})_2]^{+2}$ with Azurin, Myoglobin and Cytochrome-c." *Inorg. Chem.* **135**, 33-35.
- Coucouvanis, D., Challen, P. R., Koo, S.-M., Davis, W. M., Butler, W. and Dunham, W. R. (1989). "Stepwise Synthesis and Structural Characterization of the $[\text{MoFe}_3\text{S}_4\text{Cl}(\text{Clcat})]_2(\mu_2\text{-S})_2^{6-}$ and $[\text{MoFe}_3\text{S}_4\text{Cl}(\text{Clcat})]_2(\mu_2\text{-S})(\mu_2\text{-OH})^{5-}$ Doubly Bridged Double Cubanes Obtained by the Coupling of $[\text{MoFe}_3\text{S}_4]$ Clusters." *Inorg. Chem.* **28**, 4181-4183.
- Demadis, K. D. and Coucouvanis, D. (1995). "Synthesis, Structural Characterization, and Properties of New Single and Double Cubanes Containing the MoFe_3S_4 Structural Unit and Molybdenum-Bound Polycarboxylate Ligands. Clusters with a Molybdenum-Coordination Environment Similar to That in the Iron-Molybdenum Cofactor of Nitrogenase." *Inorg. Chem.* **34**, 436-448.
- Eaton, W. A. and Lovenberg, W. (1973). The Iron-Sulfur Complex in Rubredoxin. In *Iron Sulfur Proteins* (Editor), 131-163. Academic Press, New York.
- Evans, D. J. and Newton, M. S. (1994). "Insulation to Lattice Effects on Quadrupole Splitting in the Mossbauer-Spectra of Salts of $[\text{Fe}_4\text{S}_4(\text{Cys})_4]^{2-}$ A Ferredoxin Model." *Inorg. Chim. Acta* **217**, 3-5.
- Excoffon, P., Laugier, J. and Lamotte, B. (1991). "Influence of a Phase Transition in the Solid State on the Structure of a Synthetic Iron-Sulfur Cubane." *Inorg. Chem.* **30**, 3075-3081.
- Frey, M., Sicker, L., Payan, F., Haser, R., Bruschi, M., Pepe, G. and LeGall, J. (1987). "Rubredoxin from *Desulfovibrio gigas* A Molecular Model of the Oxidized Form at 1.4Å Resolution." *Journal of Molecular Biology* **197**, 525-541.
- Gray, H. B. (1986). "Long range electron-transfer in blue copper proteins." *Chem. Soc. Rev.* **15**, 17-30.
- Gray, H. B. and Malmström (1983). "On the Relationship between Protein-Forced Ligand Fields and the Properties of Blue Copper Centers." *Comments in Inorganic Chemistry* **2**, 203-209.
- Gray, H. B. and Solomon, E. I. (1981). Electronic Structures of Blue Copper Proteins. Copper Proteins. John Wiley & Sons, Inc. New York; pp. 1-39.
- Huang, Y.-H., Moura, I., Moura, J. J. G., LeGall, J., Park, J.-B., Adams, M. W. W. and Johnson, M. K. (1993). "Resonance Raman Studies of Nickel Tetrathiolates and Nickel-Substituted Rubredoxins and Desulfiredoxin." *Inorg. Chem.* **32**, 406-412.
- Kambayashi, H., Nagao, H., Tanaka, K., Nakamoto, M. and Peng, S. M. (1993). "Stabilization of Superoxidized Form of Synthetic Fe_4S_4 Cluster as the 1st Model of High-Potential Iron-Sulfur Proteins in Aqueous-Media." *Inorg. Chim. Acta* **209**, 143-149.
- Kim, J. and Rees, D. C. (1992). "Structural Models for the Metal Center in the Nitrogenase Molybdenum-Iron Protein." *Science* **257**, 1677-1682.
- Kodaka, M., Tomohiro, T. and Okuno, H. (1991). "Application of Electrostatic Model to Redox Potentials of Tetranuclear Iron Sulfur Clusters." *J. Phys. Chem.* **95**, 6741-6744.

Kostic, N. M., Margalit, R., Che, C. M. and Gray, H. B. (1983). "Kinetics of Long-Distance Ruthenium to Copper Electron Transfer in [Pentaammineruthenium Histidine-83]Azurin." *J. Amer. Chem. Soc.* **105**, 7765-7767.

Langen, R., Jensen, G. M., Jacob, U., Stephens, P. J. and Warshel, A. (1992). "Protein Control of Iron-Sulfur Cluster Redox Potentials." *J. Biol. Chem.* **267**, 25625-25627.

Laughlin, L. J. and Coucouvanis, D. (1995). "Use of $[\text{Mofe}_3\text{S}_4]^{3+}$ Single Cubanes in the Catalytic Reduction of Acetylene to Ethylene and Ethane. Identification of Molybdenum and Iron Atoms as Catalytic Sites during Substrate Reduction and Implications for Nitrogenase Action." *J. Am. Chem. Soc.* **117**, 3118-3125.

Lovenberg, W. and Sobel, B. E. (1965). "Rubredoxin: A new electron transfer protein from *Clostridium pasteurianum*." *Proc. Natl. Acad. Sci. USA* **54**, 193-199.

Maelia, L. E., Millar, M. and Koch, S. A. (1992). "General Synthesis of Iron(III) Tetrathiolate Complexes. Structural and Spectroscopic Models for the $[\text{Fe}(\text{Cys-S})_4]$ Center in Oxidized Rubredoxin." *Inorg. Chem.* **31**,

Malinak, S. M., Demadis, K. D. and Coucouvanis, D. (1995). "Catalytic Reduction of Hydrazine to Ammonia by the VF3S4 Cubanes. Further Evidence for the Direct Involvement of the Heterometal in the Reduction of Nitrogenase Substrates and Possible Relevance to the Vanadium Nitrogenases." *J. Am. Chem. Soc.* **117**, 3126-3133.

Marcus, R. A. and Sutin, N. (1985). "Electron Transfer in Chemistry and Biology." *Biochem. Biophys. Acta.* **811**, 265-322.

Meyer, T. E. and Cusanovich, M. A. (1989). "Structure, Function and Distribution of Soluble Bacterial Redox Proteins." *Biochem. Biophys. Acta* **975**, 1-28.

Mizoguchi, T. J., Di Bilio, A. J., Gray, H. B. and Richards, J. H. (1992). "Blue to Type 2 Binding. Copper (II) and Cobalt (II) Derivatives of a Cys112Asp Mutant of *Pseudomonas aeruginosa* Azurin." *J. Am. Chem. Soc.* **25**, 10076-10078.

Ohno, R., Ueyama, N. and Nakamura, A. (1991). "Influence of the Distal Para Substituent through NH—S Hydrogen Bonds on the Positive Shift of the Reduction Potentials of $[\text{Fe}_4\text{S}_4(\text{Z-cys-Gly-NHC}_6\text{H}_4\text{-p-X})_4]^{2-}$ (X= H, Ome, F, Cl, CN) Complexes." *Inorg. Chem.* **30**, 4887-4891.

Roth, E. K. H. and Jordanov, J. (1992). "Oxidation Reactions of $[\text{Fe}_4\text{S}_4(\text{S}2,4,6\text{-}(i\text{-Pr})_3\text{C}_6\text{H}_2)_4]^{2-}$ and an Oxidative Conversion of the Fe_4S_4 Core into an Fe_3S_4 Center." *Inorg. Chem.* **31**, 240-243.

Sun, W.-Y., Ueyama, N. and Nakamura, A. (1991). "Reduced Rubredoxin Models Containing Z-Cys-Pro-Leu-Cys-Gly-NH-C₆H₄-p-X (X= MeO, H, F, CN); Electronic Influence by a Distant Para Substituent through NH—S Hydrogen Bonds." *Inorg. Chem.* **30**, 4026-4031.

Sun, W.-Y., Ueyama, N. and Nakamura, A. (1993). "An Electronic Influence of a Distant Aromatic Ring in Reduced Rubredoxin Models. Iron(II) Complexes with Z-Cys-Pro-Leu-Cys-Glu-X (X= NHCH₂C₆H₄-p-F, NHCH₂CH₂C₆H₄-p-F, and Phe-OMe)." *Inorg. Chem.* **32**, 1095-1100.

Ueyama, N., Oku, H. and Nakamura, A. (1992). "2-Electron-Transfer Oxidation by Polymer-Supported $[\text{Fe}_4\text{S}_4(\text{sr})_4]^{2-}$ Complex." *J. Mol. Catal.* **74**, 451-458.

Ueyama, N., Sun, W.-Y. and Nakamura, A. (1992). "Evidence for Intramolecular NH—S Hydrogen Bonds from ^2H -NMR Spectroscopy of Reduced-Rubredoxin Model Fe(II) Complexes with Bidentate Peptide Ligands." *Inorg. Chem.* **31**, 4063-4059.

Walters, M. A., Dewan, J. C., Min, C. and Pinto, S. (1991). "Models of Amide-Cysteine Hydrogen Bonding in Rubredoxin: Hydrogen Bonding between Amide and Benzenethiolate in $[(\text{CH}_3)_3\text{NCH}_2\text{CONH}_2]_2[\text{Co}(\text{SC}_6\text{H}_5)_4]^{+1/2}\text{CH}_3\text{CN}$ and $[(\text{CH}_3)_3\text{NCH}_2\text{CONH}_2][\text{SC}_6\text{H}_5]$." *Inorg. Chem.* **30**, 2656-2662.

Werth, M. T., Kurtz, D. M., Jr., Howes, B. D. and Huynh, B. H. (1989). "Observation of $\text{S}=\text{2}$ EPR Signals from Ferrous Iron-Tiolate Complexes. Relevance to Rubredoxin-Type Sites in Proteins." *Inorg. Chem.* **28**, 1357-1361.

Winkler, J. R. and Gray, H. B. (1992). "Electron-Transfer in Ruthenium-Modified Proteins." *Chem. Rev.* **92**, 369-379.

Yanada, K., Nagano, T. and Hirobe, M. (1993). "Comparison of The Reactivities of $[\text{Fe}_4\text{S}_4(\text{SPh})_4]^{2-}$ and $[\text{Fe}_2\text{S}_2(\text{SPh})_4]^{2-}$." *Chemical & Pharmaceutical Bulletin* **41**, 208-210.

Zhou, J., Scott, M. J., Hu, Z. G., Peng, G., Munck, E. and Holm, R. H. (1992). "Synthesis and Comparative Reactivity and Electronic Structural Features of $[\text{MFe}_3\text{S}_4]^{z+}$ Cubane-Type Clusters ($\text{M} = \text{Fe}, \text{Co}, \text{Ni}$)." *J. Am. Chem. Soc.* **114**, 10843-10854.

Chapter 1
Rubredoxin from
Pyrococcus furiosus

1.1 Introduction to Rubredoxin

In the early 1980's organisms were discovered that thrived at temperatures near 100 °C. They are termed hyperthermophiles due to their optimal growth temperatures between 80 and 110 °C and lack of growth at temperatures below 60 °C (Huber, Stoffers et al. 1990). These organisms have been primarily classified as belonging to the family *Archaea* (Woese, Kandler et al. 1990) (formerly *Archaeobacterium*) and are found in both terrestrial and marine geothermal environments. The hyperthermophiles (from the domain *Archaea*) are believed to be the most ancient organisms in their domain and the two hyperthermophilic organisms of the domain *Bacteria* appear to be the most ancient in that domain as well. This observation leads one to the postulate that these organisms are the ancestors of all life and that life at lower temperature is the result of evolutionary pressure to adapt to temperatures significantly lower than 100 °C.

The domain *Archaea* can be divided into two kingdoms: *Euryarchaeota* (containing the methanogens and their relatives) and *Crenarchaeota* (made up of the hyperthermophiles) (Woese, Kandler et al. 1990). The majority of the hyperthermophiles are strictly anaerobic heterotrophs that can grow on complex mixtures of peptides and depend on the reduction of elemental sulfur (S^0) to H_2S to maintain growth (Adams 1993). This general dependence on the reduction of sulfur for significant growth has limited the study of the physiology of hyperthermophiles to those few species that will grow in the absence of S^0 . The organism *Pyrococcus furiosus* is one such species and the remainder of this discussion will focus mainly on this organism and primarily on the structure of the 53 amino acid, iron-sulfur protein, rubredoxin, isolated from it.

A novel genus of thermophilic marine organism growing between 70 and 103 °C was first isolated by Stetter and Fiala (Fiala and Stetter 1986) from sediments surrounding a geothermal vent at the beach of Porto di Levante, Vulcano, Italy. The new genus was named *Pyrococcus*, meaning the "fireball." The species and strain is *Pyrococcus furiosus*

Vc 1 (DSM 3638). *Pyrococcus furiosus* grows optimally at 100 °C with a doubling time of 37 min. The organism grows heterotrophically in either the presence or absence of sulfur (S^0) by the fermentation of either amino acids or oligosaccharides where the fermentation products are CO_2 , acetate, acetoin and H_2 . In the presence of elemental sulfur, the latter, which inhibits growth, is converted to H_2S by hydrogenase and thus removed from the environment. The pathway for the reduction of elemental sulfur by H_2 is not understood at this time. Enzymes have been isolated in *Pyrococcus furiosus* that are responsible for the metabolism of starch, glucose and pyruvate as well as enzymes with proteolytic activity. See Figure 1.1-1 for a simplified schematic of the metabolic pathways in *P. furiosus*.

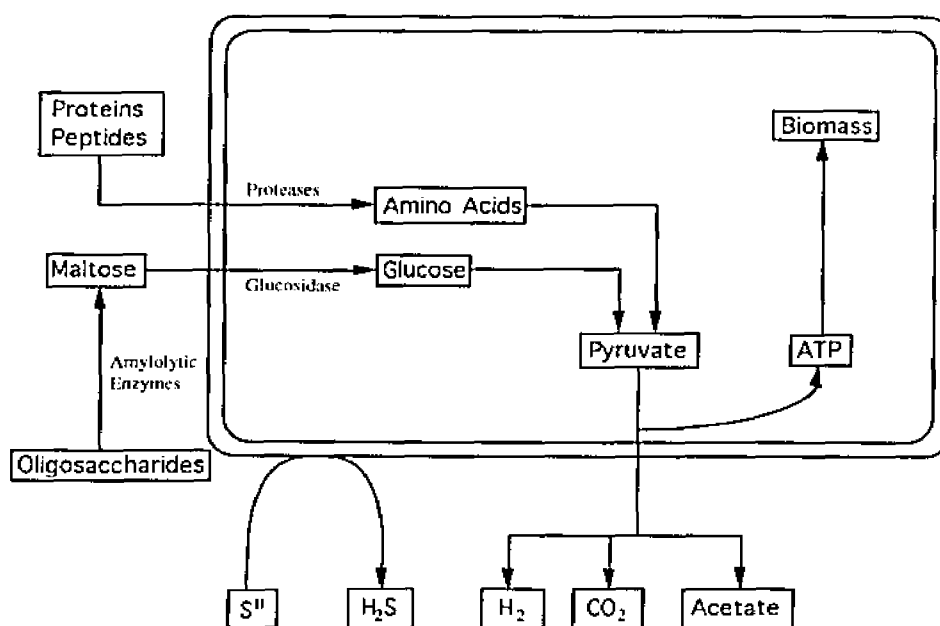


Figure 1.1-1. A schematic representation of the metabolic pathways in *P. furiosus* illustrating the extracellular cleavage of oligosaccharides and the intercellular proteolysis of peptides and proteins and the hydrolysis of maltose.

P. furiosus grows well on complex organic substrates like yeast extract or tryptone. Whole proteins or enzymatically hydrolyzed casein will support sustained growth, but the organism will not grow on single amino acids. Experiments have shown, however, that *P. furiosus* will incorporate cysteine and methionine (Blumentals, Itoh et al. 1990) and the addition of single amino acids to the media can significantly increase growth (Kelly,

Blumentals et al. 1992). This suggests that *P. furiosus* may be auxotrophic for one or more amino acids. It is reasonable to assume that the peptide requirement for sustained growth may reflect the greater stability, with respect to temperature, of peptides over single amino acids.

Carbohydrates are not required for growth. *P. furiosus* will grow on maltose and starch but will not grow on glucose. Enzymes responsible for the hydrolysis of starch (Constantino, Brown et al. 1990) and the metabolism of glucose (Mukund and Adams 1991) have been isolated so it would seem that *P. furiosus* is not able to transport glucose across the membrane but is able to metabolize glucose once maltose or starch have been hydrolyzed inside the cell.

The pathway for the metabolism of sugars by *P. furiosus* is a subject that has attracted much interest recently. Initial investigations suggested that *P. furiosus* glucose metabolism proceeded by a unique pathway. It has been demonstrated that many archaebacteria utilized the Entner-Doudoroff pathway rather than the Embden-Meyerhof pathway observed in most all higher organisms. The main difference arises from the lack of a 6-phosphofructosekinase (Conway 1992; Danson and Hough 1992). This enzyme is present in eukaryotes and many anaerobic eubacteria where the Embden-Meyrhof glucolytic pathway is operative. This pathway produces two molecules of ATP per molecule of metabolized glucose. Many strictly aerobic organisms do not contain this enzyme and therefore use the Entner-Doudoroff pathway where only one molecule of ATP is generated per glucose. However, ATP-phosphofructosekinase has not been detected in any species of *Archaea* so it has been proposed that variations of the Entner-Doudoroff pathway are operative here. *P. furiosus* has been proposed to metabolize glucose by a non-phosphorylated Entner-Doudoroff pathway where there is no net yield of ATP (Schäfer and Schönheit 1992).

The fate of the C₁-carbon is different in the two pathways (see Figure 1.1-2.)

(Conway 1992; Danson and Hough 1992). During Entner-Doudoroff metabolism the C₁-carbon of glucose is evolved as CO₂. The C₁-carbon of glucose becomes the methyl group of pyruvate during Embden-Meyerhof metabolism. In this case, labeled CO₂ would not be seen until much later, after pyruvate has been completely metabolized during the tricarboxylic acid cycle. This provides a convenient method for the elucidation of which cycle is responsible for the major portion of glucose metabolism. All of the enzymes necessary for gluconeogenesis via the Embden-Meyerhof pathway are present in *P. furiosus* (Schäfer, Xavier et al. 1994). Cell free extracts of *P. furiosus* contain all the enzymes necessary for the modified non-phosphorylated Entner-Doudoroff pathway except gluconate dehydratase (Schäfer and Schönheit 1992).

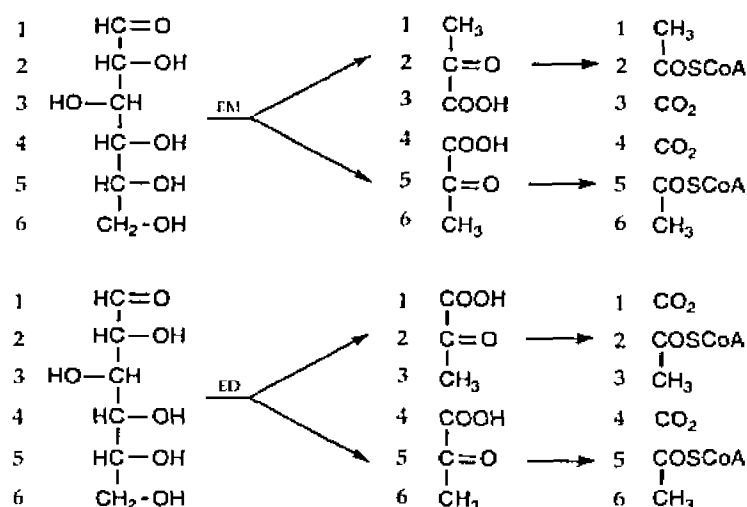


Figure 1.1-2 Labeling patterns for glycolysis. CO₂ from the Embden-Meyerhof (EM) pathway will be formed from the C₃ and C₄ carbons. The Entner-Doudoroff (ED) pathway will form CO₂ from the C₁ and C₄ carbons.

Kengen et al. used *in vivo* ¹³C NMR and enzyme measurement in cell free extracts to trace the fate of labeled glucose during metabolism by *P. furiosus*. Their findings indicate that the Embden-Meyerhof pathway is the major glycolytic pathway and it is operative by way of a number of novel ADP-dependent kinases (Kengen, de Bok et al. 1994). It could be that *P. furiosus* would gain an advantage from such an arrangement on the basis that ATP would be less stable than ADP at optimal growth temperatures and that

ADP would supply the same amount for energy. Similar ^{13}C NMR experiments with labeled glucose by Schäfer on cell free extracts suggest a novel pathway utilizing a combination of reactions from both the Embden-Meyerhof and the non-phosphorylated Entner-Doudoroff pathways (Schäfer, Xavier et al. 1994).

As mentioned above, the reduction of elemental sulfur (S^0) to H_2S is a process common to all hyperthermophiles. Most hyperthermophilic archaea appear to utilize sulfur for respiration and are therefore obligately dependent on its presence for growth. Experiments have shown that the addition of sulfur to the culture medium stimulates the growth of *P. furiosus* (Schicho, Ma et al. 1993). Although *P. furiosus* is not obligately dependent on the presence of sulfur for growth, it has been proposed that sulfur reduction to H_2S is a means of removing H_2 from the environment (H_2 being a growth inhibitor). The means by which this process occurs has generated much interest.

Elemental sulfur has many allotropes and the form that is most thermodynamically stable at the optimal growth temperature of *P. furiosus* is monoclinic β -sulfur. This form of sulfur is an eight membered ring that is essentially insoluble in water. The question then becomes: how does the organism sequester this insoluble form of sulfur for reduction to H_2S . The two are possibilities (1) the organism must be in direct contact with the solid sulfur substrate or (2) the organism must, in some way, convert the sulfur compound into a soluble form. Experiments have shown that *P. furiosus* does not need to be in direct contact with elemental sulfur and that soluble polysulfides are a likely candidate for the soluble sulfur substrate reduced by H_2 evolved by the organism (Blumentals, Itoh et al. 1990). Polysulfides can be formed by nucleophilic attack on β -sulfur in ionic solutions at around 100°C . It is, therefore, reasonable to expect polysulfides to exist in the natural environment of *P. furiosus*. It is interesting to note that S^0 reduction to H_2S also occurs abiotically at the growth temperature of *P. furiosus*, so it is not entirely clear as to the organism's role in sulfur reduction.

In their experiments on the effect of S^0 added to an energy-limited culture system, Schicho et al. found that the maximum yield of cells nearly doubled when the organism was grown in the presence of sulfur. This suggested that the reduction of sulfur might be an energy conserving reaction (Schicho, Ma et al. 1993). In their pursuit of the enzymes responsible for the sulfur reduction, they found that the hydrogenase protein of *P. furiosus* was found to have a S^0 -reducing activity (Ma, Schicho et al. 1993). It was further found that reduction of sulfur to H_2S appears to be a common reaction catalysed by hydrogenases from not only hyperthermophiles but some mesophiles as well. Interestingly, it was discovered that at pH 7.6 the reduction of sulfur by hydrogenase in *P. furiosus* proceeded at ≈ 3 times the rate in the presence of the iron-sulfur redox protein rubredoxin as it did in the absence of rubredoxin. Rubredoxin had no effect at pH 8.4 and did not stimulate H_2 oxidation or H_2 evolution in the pH range 7.0-8.4. The ferredoxin of *P. furiosus*, which is believed to be the *in vivo* electron donor to hydrogenase for H_2 activity, had no effect on S^0 -reduction. A physiological role for rubredoxin has never been established in any of the organisms in which it has been found. Therefore, the finding that rubredoxin catalyzes the reduction of sulfur by hydrogenase at physiological pH presents the intriguing possibility that this might be rubredoxin's role in *P. furiosus*.

Because *P. furiosus* grows at such high temperatures, structural investigations of the proteins isolated from this organism present the opportunity to study the determinants of protein stability at elevated temperatures. An understanding of the factors that affect the thermostability of proteins has obvious consequences in protein engineering for industrial applications. Rubredoxin presents a unique opportunity in this respect because the x-ray structures of rubredoxins from four mesophilic organisms have been determined to near atomic resolution (i.e. between 1.2 and 1.5 Å). These four other rubredoxins are structurally similar and show no exceptional thermostability. Therefore, the expectation is that a similar structural analysis of the rubredoxin from *P. furiosus*, which has been

shown to be stable to at least 110 °C (Klump, Adams et al. 1994), will provide insights into the factors that lend thermostability to proteins. The determination and analysis of the structure of rubredoxin with respect to its thermostability will be the focus for the remainder of this chapter.

		5	10	15	20	25	30	35	40	45	50																																												
		↓	↓	↓		↓		↓			↓																																												
RdPf		A	K	W	V	C	K	I	C	G	Y	I	D	E	D	A	G	D	P	D	N	G	I	S	P	G	T	K	F	E	E	L	P	D	D	W	V	C	P	I	C	G	A	P	K	S	E	F	E	K	L	E	D		
RdDg		M	D	I	Y	V	C	T	V	C	G	Y	E	Y	D	P	A	K	G	D	P	D	S	G	I	K	P	G	T	K	F	E	D	L	P	D	D	W	A	C	P	V	C	G	A	S	K	D	A	F	E	K	Q		
RdDv		M	K	K	Y	V	C	T	V	C	G	Y	E	Y	D	P	A	E	G	D	P	D	N	G	V	K	P	G	T	S	F	D	D	L	P	A	D	W	V	C	P	V	C	G	A	P	K	S	E	F	E	A	A		
RdCp		M	K	K	Y	T	C	T	V	C	G	Y	I	Y	D	P	E	D	G	D	P	D	G	V	N	P	G	T	D	F	K	D	I	P	D	D	W	V	C	P	L	C	G	V	G	K	D	E	F	E	E	V	E	E	
RdDd		M	Q	K	Y	V	C	N	V	C	G	Y	E	Y	D	P	A	E	H	D	-----	N	V	P	F	D	Q	L	P	S	S	W	C	C	P	V	C	G	V	S	K	D	Q	F	S	P	A								
Cpf		M	K	K	F	I	C	D	V	C	G	Y	I	Y	D	P	A	V	G	D	P	D	N	G	V	E	P	G	T	E	F	K	D	I	P	D	D	W	V	C	P	L	C	G	V	D	K	S	Q	F	S	E	T	E	E
Cts		M	E	K	W	Q	C	T	V	C	G	Y	I	Y	D	P	E	V	G	D	P	T	Q	N	I	P	P	G	T	K	F	E	D	L	P	D	D	W	V	C	P	D	C	G	V	D	K	D	Q	F	E	K	I		
Cbt		M	Q	K	Y	V	C	S	V	C	G	Y	V	Y	D	P	A	D	G	E	P	D	D	I	D	P	G	T	G	F	E	D	L	P	D	E	W	V	C	P	V	C	G	V	D	K	D	L	F	E	P	E	S		
DvM		M	K	K	Y	V	C	T	V	C	G	Y	E	Y	D	P	A	E	G	D	P	D	N	G	V	K	P	G	T	A	F	E	D	V	P	A	D	W	A	C	P	V	C	G	A	P	K	S	E	F	E	P	A		
DvH		M	K	K	Y	V	C	T	V	C	G	Y	E	Y	D	P	Q	E	G	D	P	D	N	G	V	K	P	G	T	S	F	D	D	L	P	A	D	W	V	C	P	V	C	G	A	P	K	S	E	F	E	A	A		
Me		M	D	K	Y	E	C	S	I	C	G	Y	I	Y	D	E	A	E	G	D	-	D	G	N	V	A	A	G	T	K	F	A	D	L	P	A	D	W	V	C	P	T	C	G	A	D	K	D	A	F	V	K	M	D	
Pa		M	Q	K	F	E	C	T	L	C	G	Y	I	Y	D	P	A	L	V	G	P	D	T	P	D	Q	D	G	-	A	F	E	D	V	S	E	N	W	V	C	P	L	C	G	A	G	K	E	D	F	E	V	Y	E	D
Bm		M	Q	K	Y	V	C	D	I	C	G	Y	V	Y	D	P	A	V	G	D	P	D	N	G	V	A	P	G	T	A	F	A	D	L	P	E	D	W	V	C	P	E	C	G	V	S	K	D	E	F	S	P	E	A	

Figure 1.1-3. Sequence alignment for the rubredoxins from anaerobic bacterium. The top five sequences are for the rubredoxins whose x-ray structures have been solved. The conserved cysteines are given in bold face. The symbols ↓ and | indicate residues that are unique to RdPf and residues that are strictly conserved in all known rubredoxins, respectively. The abbreviations and references are: RdPf, *P. furiosus* (Blake, Park et al. 1991); RdDg, *D. gigas* (Bruschi 1976a); RdDv, *D. vulgaris* (Bruschi 1976b); RdCp, *C. pasteurianum* (Watenpaugh, Seiker et al. 1973; Yasunobo and Tanaka 1983); RdDd, *D. desulfuricans* (Hormel, Walsh et al. 1986); Cpf, *C. perfringens* (Seki, Seki et al. 1989); Cts, *C. thermosaccharolyticum* (Meyer, Gagnon et al. 1990); Cbt *C. Thiosulfatophilum* (Woolley and Meyer 1987); DvM, *D. vulgaris* strain Miyazaki (Shimizu, Ogata et al. 1989); DvH, *D. vulgaris* strain Hildenborough (Bruschi 1976b; Voordouw 1988); Me, *Megasphaera elsdenii* (Bachmeyer, Yasunobo et al. 1968b); Pa, *Peptococcus aerogenes* (Bachmeyer, Benson et al. 1968a); Bm, *Butyribacterium methyltropicum* (Saeki, Yao et al. 1989).

References

- Adams, M. W. W. (1993). "Enzymes and proteins from organisms that grow near and above 100°C." *Ann. Rev. Microbiol.* **47**, 627-658.
- Bachmeyer, H., Benson, A. M., Yasunobo, K. T., Garrad, W. T., et al. (1968a). "Nonheme Iron Proteins. IV. Structural Studies of *Micrococcus aerogenes* Rubredoxin." *Biochemistry* **7**, 986-996.
- Bachmeyer, H., Yasunobo, K. T., Peel, J. L. and Mayhew, S. (1968b). "Non-heme iron proteins." *J. Biol. Chem.* **243**, 1022-1030.

Blake, P. R., Park, J.-B., Bryant, F. O., Aono, S., et al. (1991). "Determinants of Protein Hyperthermostability: Purification and Amino Acid Sequence of Rubredoxin from the Hyperthermophilic Archaeobacterium *Pyrococcus furiosus* and Secondary Structure of the Zinc Adduct by NMR." *Biochemistry* **30**, 10885-10895.

Blumentals, I. I., Itoh, M., Olson, G. J. and Kelly, R. M. (1990). "Role of Polysulfides in Reduction of Elemental Sulfur by the Hyperthermophilic Archaeobacterium *Pyrococcus furiosus*." *Appl. Environ. Microbiol.* **56**, 1255-1262.

Bruschi, M. (1976a). "The amino acid sequence of the rubredoxin from the sulfate reducing bacterium, *Desulfovibrio gigas*." *Biochem. Biophys. Res. Commun.* **70**, 615-621.

Bruschi, M. (1976b). "Non-heme iron proteins. The amino acid sequence of rubredoxin from *Desulfovibrio vulgaris*." *Biochem. Biophys. Acta* **434**, 4-17.

Constantino, H. R., Brown, S. H. and Kelly, R. M. (1990). "Purification and Characterization of an α -Glucosidase from a Hyperthermophilic Archaeobacterium, *Pyrococcus furiosus*, Exhibiting a Temperature Optimum of 105 to 115°C." *J. Bacteriol.* **172**, 3654-3660.

Conway, T. (1992). "The Entner-Doudoroff pathway: history, physiology and molecular biology." *FEMS Microbiological Reviews* **103**, 1-28.

Danson, M. J. and Hough, D. W. (1992). "The enzymology of archaeobacterial pathways of central metabolism." *Biochem. Soc. Sym.* **58**, 7-21.

Fiala, G. and Stetter, K. O. (1986). "*Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaeobacteria growing optimally at 100°C." *Arch Microbiol* **145**, 56-61.

Hormel, S., Walsh, K. A., Prickril, B. C., Titani, K., et al. (1986). "Amino-Acid Sequence of Rubredoxin from *Desulfovibrio desulfuricans* Strain 27774." *FEBS-Letters* **201**, 147-150.

Huber, R., Stoffers, P., Cheminee, J. L., Richnow, H. H., et al. (1990). "Hyperthermophilic archaeobacteria within the crater and open-sea plume of erupting Macdonald Seamount." *Nature* **345**, 179-182.

Kelly, R. M., Blumentals, I. I., Snowden, L. J. and Adams, M. W. W. (1992). "Physiological and biochemical characteristics of *Pyrococcus furiosus*, a hyperthermophilic archaeobacterium." *Annals Of The New York Academy Of Sciences* **665**, 309-319.

Kengen, S. W. M., de Bok, F. A. M., van Loo, N.-D., Dijkema, C., et al. (1994). "Evidence for the Operation of a Novel Embden-Meyerhof Pathway That Involves ADP-dependent Kinases during Sugar Fermentation by *Pyrococcus furiosus*." *J. Biol. Chem.* **269**, 17537-17541.

Klump, H. H., Adams, M. W. W. and Robb, F. T. (1994). "Life in the pressure cooker: The thermal unfolding of proteins from hyperthermophiles." *Pure & Appl. Chem.* **66**, 485-489.

Ma, K., Schicho, R. N., Kelly, R. M. and Adams, M. W. W. (1993). "Hydrogenase of the hyperthermophile *Pyrococcus furiosus* is an elemental sulfur reductase or sulfhydrogenase - evidence for a sulfur-reducing hydrogenase ancestor." *Proc. Natl. Acad. Sci. USA* **90**, 5341-5344.

Meyer, J., Gagnon, J., Sieker, L. C., van Dorsselaer, A., et al. (1990). "Rubredoxin from *Clostridium thermosaccharolyticum*: Amino Acid Sequence, Mass-Spectrometric and Preliminary Crystallographic Data." *Biochemical Journal* **271**, 839-841.

Mukund, S. and Adams, M. W. W. (1991). "The Novel Tungsten-Iron-Sulfur Protein of the Hyperthermophilic Archaeobacterium, *Pyrococcus furiosus*, Is an Aldehyde Ferredoxin Oxidoreductase." *J. Biol. Chem.* **266**, 14208-14216.

- Saeki, K., Yao, Y., Wakabayashi, S., Shen, G. J., et al. (1989). "Ferredoxin and Rubredoxin from *Butyrivibacterium methylotrophicum*: Complete Primary Structures and Construction of Phylogenetic Trees." *Journal of Biochemistry* **106**, 656-662.
- Schäfer, T. and Schönheit, P. (1992). "Maltose fermentation to acetate, CO_2 and H_2 in the anaerobic hyperthermophilic archaeon *Pyrococcus furiosus* - evidence for the operation of a novel sugar fermentation pathway." *Archives Of Microbiology* **158**, 188-202.
- Schäfer, T., Xavier, K. B., Santos, H. and Schönheit, P. (1994). "Glucose fermentation to acetate and alanine in resting cell suspensions of *Pyrococcus furiosus*: Proposal of a novel glycolytic pathway based on ^{13}C labelling data and enzyme activities." *FEMS Microbiology Letters* **121**, 107-114.
- Schicho, R. N., Ma, K., Adams, M. W. W. and Kelly, R. M. (1993). "Bioenergetics of Sulfur Reduction in the Hyperthermophilic Archaeon *Pyrococcus furiosus*." *J. Bacteriol.* **175**, 1823-1830.
- Seki, Y., Seki, S., Satoh, M., Ikeda, A., et al. (1989). "Rubredoxin from *Clostridium perfringens* - Complete Amino-Acid Sequence and Participation in Nitrate Reduction." *Journal of Biochemistry* **106**, 336-341.
- Shimizu, F., Ogata, M., Yagi, T., Wakabayashi, S., et al. (1989). "Amino-Acid Sequence and Function of Rubredoxin from *Desulfovibrio vulgaris* Miyazaki." *Biochemie* **71**, 1171-1177.
- Voordouw, G. (1988). "Cloning of genes encoding redox proteins of known amino sequence from a library of the *Desulfovibrio vulgaris* (Hildenborough) genome." *Gene* **67**, 75-83.
- Watenpaugh, K. D., Seiker, L. C., Herriot, J. R. and Jensen, L. H. (1973). "Refinement of the model of a protein: Rubredoxin at 1.5Å resolution." *Acta Crystallogr.* **B29**, 943.
- Woese, C. R., Kandler, O. and Wheelis, M. L. (1990). "Towards a natural system of organisms: Proposal for the domains Archae, Bacteria, and Eucarya." *Proc. Natl. Acad. Sci. USA* **87**, 4576-4579.
- Woolley, K. J. and Meyer, T. E. (1987). "The complete amino-acid sequence of rubredoxin from the green photrophic bacterium *Chlorobium thiosulfatophilum* strain PM." *Eur. J. Biochem.* **163**, 161-166.
- Yasunobo, K. T. and Tanaka, M. (1983). The type, distribution in nature, structure-function, and evolutionary data of the iron-sulfur proteins. In *Iron-Sulfur Proteins*. (W. Lovenberg, eds.) pp. 27-130. Academic Pres, New York.

1.2 Experimental

1.2.1 Crystallization

The growth of *Pyrococcus furiosus* and the isolation of rubredoxin (RdPf) were as previously described (Bryant and Adams 1989; Blake, Park et al. 1991). Crystals of the oxidized form of rubredoxin were grown using the hanging drop method by equilibrating a 4 μ l drop against a solution containing 30% 2-methyl-2,4-pentanediol, 2.2M (NH₄)₂SO₄ and 0.1M Tris/HCl pH 8.5. The drop contained 2 μ l of protein solution (~60 mg/ml protein, 0.3M NaCl and 50mM Tris/Cl pH 8.0) and 2 μ l of buffered ammonium sulfate solution (3.2M (NH₄)₂SO₄ and 0.15M Tris/Cl pH 8.5). Deep red rectangular shaped crystals of approximate dimensions 0.3 x 0.3 x 0.4 mm grew overnight in the orthorhombic space group $P 2_1 2_1 2_1$ with unit cell dimensions of $a=34.6\text{\AA}$, $b=35.5\text{\AA}$, $c=44.4\text{\AA}$ and $V=54,536\text{\AA}^3$. Assuming 4 molecules in the unit cell, the ratio of volume to molecular weight (Matthews 1968), V_m , is calculated to be $2.30\text{\AA}^3/\text{dalton}$. At low temperature (-161°C) the unit cell dimensions each shrink by approximately 2.5% to $a=33.8\text{\AA}$, $b=34.6\text{\AA}$, $c=43.4\text{\AA}$ with a corresponding 7% loss in volume to $V=50,755\text{\AA}^3$.

Crystals of the reduced form of rubredoxin were obtained by adding a minimum amount (several grains) of sodium dithionite (Na₂S₂O₄) to drops containing crystals of the oxidized form of rubredoxin. Upon addition of sodium dithionite, the crystals gradually lose their deep red color over a period of one to two minutes, becoming colorless and extremely fragile. When the crystals became colorless, they were removed from the drop and mounted on a glass spatula for data collection and immediately placed into a -161°C stream of N₂. The space group and measured unit cell dimensions are virtually identical to those of the oxidized form ($P 2_1 2_1 2_1$, $a=33.8\text{\AA}$, $b=34.5\text{\AA}$, $c=43.2\text{\AA}$ and $V=50,375\text{\AA}^3$).

References

Blake, P. R., Park, J.-B., Bryant, F. O., Aono, S., et al. (1991). "Determinants of Protein Hyperthermostability: Purification and Amino Acid Sequence of Rubredoxin from the Hyperthermophilic Archaeobacterium *Pyrococcus furiosus* and Secondary Structure of the Zinc Adduct by NMR." *Biochemistry* 30, 10885-10895.

Bryant, F. O. and Adams, M. W. W. (1989). "Characterization of Hydrogenase from the Hyperthermophilic Archaeobacterium, *Pyrococcus furiosus*." *J. Biol. Chem.* **264**, 5070-5079.

Matthews, B. W. (1968). "Solvent Content of Protein Crystals." *Journal of Molecular Biology* **33**, 491-497.

1.2.2 Data Collection

All intensity data were collected on a Siemens X-1000 area detector with the crystal to detector distance set at 8.0 cm. Cu K α ($\lambda=1.5418\text{\AA}$) radiation was generated by a Siemens rotating anode generator operating at 4.5 kW (50 kV x 90 mA) with a graphite monochromator. For low temperature data collection, the crystals were cooled to -161°C with a stream of cold nitrogen generated by a Siemens LT-2A low temperature device, employing the cryocrystallographic techniques of Hope (Hope 1988; Hope 1990).

Data for initial phasing and model building were collected at room temperature to a maximum resolution of 2.2\AA with the detector center set at $2\theta=10^\circ$. Unfortunately, the crystal dissolved during data collection and it was not possible to obtain data from a second orientation in Φ . A total of 7767 observations with an average $\frac{1}{\sigma I} = 35$ were integrated using XENGEN (Howard, Gilliland et al. 1987). Only 5067 of the integrated observations could be successfully scaled using the program package ROCKS (Bethge 1984; Reecke 1984). The resulting data set (2322 reflections) was 66.9% complete to 2.2\AA with $R_{\text{merge}}=8.8\%$ (see Table 1.2.2-1). A Wilson plot (Wilson 1942) for the data indicates an average $B=16.2\text{\AA}^2$.

Table 1.2.2-1. Data collection statistics for the four data sets used in the refinement of RdPf.

Data set	Resolution range	Unique reflections	% Observed	Number of observations	R_{merge}
R. T. ^a	∞ - 2.2\AA	2322	66.9	7767	0.088
-161°C (ox) ^b	∞ - 1.1\AA	17366	81.2	78073	0.044
$2\theta=20^\circ$ ^c	∞ - 1.8\AA	4439	87.3	18963	0.025
-161°C (red) ^d	∞ - 1.5\AA	7859	90.9 ^e	45990	0.037

^aroom temperature, oxidized RdPf data set used for the initial molecular replacement calculations

^bcomplete oxidized RdPf data set collected at -161°C , merging data collected with the detector center at both $2\theta=20^\circ$ and $2\theta=60^\circ$

^clow temperature, oxidized RdPf data collected with the detector center set at $2\theta=20^\circ$ and used for structure refinement

^dlow temperature, reduced RdPf data collected with the detector center set at $2\theta=30^\circ$ and used for structure refinement

^e100% complete to 1.8\AA resolution

Data used for the refinement of the structure were collected at -161°C from a single crystal of the oxidized form to a maximum resolution of 1.1\AA , employing two orientations of the area detector. Data to 1.8\AA resolution were collected with the detector centered at $2\theta=20^{\circ}$ with two orientations around the Φ axis. Data between 2.8\AA and 1.1\AA resolution were collected with the detector at $2\theta=60^{\circ}$ using three different rotations around the Φ axis (see Table 1.2.2-2). The 78023 observations (average $\frac{1}{\sigma I} = 24$) obtained from the seven different orientations were integrated, scaled and merged using XENGEN resulting in a data set (17366 reflections) 81.2% complete to 1.1\AA with $R_{\text{merge}}=4.4\%$ (see Table 1.2.2-1) and a Wilson $B = 3.9\text{\AA}^2$.

Table 1.2.2-2 Data was collected on crystals of approximately $0.3 \times 0.3 \times 0.3 \text{ mm}$ in size using the following orientations, scanning in Ω at the 2θ and Φ angles shown with a crystal to detector distance of 8.0 cm .

Oxidized						
Orientation #	2θ	Starting Φ	Starting Ω	Frame Width ($^{\circ}$)	Time (sec)	# Frames
1	60°	120°	30°	0.10	150	1000
2	60°	40°	30°	0.10	150	1000
3	20°	120°	30°	0.10	150	352
4	20°	120°	10°	0.15	75	1233
5	60°	120°	75°	0.15	150	350
6	20°	30°	35°	0.15	75	500
7	60°	75°	75°	0.15	150	343
Reduced						
1	30°	80°	45°	0.15	75	1200
2	30°	170°	45°	0.15	75	1200
3	30°	125°	45°	0.15	75	1200

Data were collected at -161°C from a single crystal of the reduced form to a maximum resolution of 1.5\AA with a single setting of the detector at $2\theta=30^{\circ}$ and in three orientations around the Φ axis (see Table 1.2.2-2). Data (45990 observations, average $\frac{1}{\sigma I} = 23$) were integrated, scaled and merged as above. The final data set (Wilson $B = 7.8\text{\AA}^2$) containing 7859 reflections is 98.7% complete to 1.6\AA and 90.9% complete to

1.5Å with $R_{\text{merge}}=3.7\%$.

References

Bethge, P. (1984). "VAX adaptation of the ROCKS crystallographic program system." *J. Appl. Crystallogr.* **17**, 215.

Hope, H. (1988). "Cryocrystallography of Biological Macromolecules: a Generally Applicable Method." *Acta Crystallographica B* **44**, 22-26.

Hope, H. (1990). "Crystallography of Biological Macromolecules At Ultra-Low Temperature." *Annual Review of Biophysics and Biophysical Chemistry* **19**, 107-126.

Howard, A. J., Gilliland, G. L., Finzel, B. C., Poulos, T. L., et al. (1987). "The Use of an Imaging Proportional Counter in Macromolecular Crystallography." *Journal of Applied Crystallography* **20**, 383-387.

Reecke, G. N., Jr. (1984). "The ROCKS System of Computer Programs for Macromolecular Crystallography." *Journal Applied Crystallography* **17**, 125-130.

Wilson, A. J. C. (1942). "Determination of Absolute from Relative X-Ray Intensity Data." *Nature* **150**, 151-152.

1.2.3 Structure Determination

Molecular Replacement

A model was prepared from RdCp by truncating side chains that were not common to each structure to alanine. This resulted in 338 atoms common to both RdPf and RdCp. The origin of the coordinate system was then adjusted to coincide with the geometric center of the molecule.

The structure of oxidized RdPf was solved by molecular replacement, using as the search model 337 atoms (excluding Fe) common to RdPf and RdCp. Rotation functions were calculated with the Fast Rotation Function of Crowther (Crowther 1972), using all measured reflections between 8-3Å and integration radii varying between 12 and 20Å. A peak consistently appeared in the $\beta=75^\circ$ section corresponding to the Euler angles $\theta_1=125^\circ$, $\theta_2=75^\circ$ and $\theta_3=-40^\circ$. The oriented molecule was positioned in the cell by a brute force translation search using reflections between 6.5 and 3.5Å resolution. A solution was found with a correlation coefficient of 0.54 that was 50% greater than the next highest peak. This solution required a translation by $x=0.365$ (12.63Å), $y=0.067$ (2.40Å) and $z=0.122$ (5.46Å). The R-factor ($R = \sum ||F_O| - |F_C|| / \sum |F_O|$) for the rotated and translated model (calculated with the refinement package TNT (Tronrud, Ten Eyck et al. 1987)) was $R=0.45$.

Patterson Results

The availability of high resolution data offered an opportunity to verify the correctness of the molecular replacement solution. This was done after the structure had been solved and refined. A Patterson map was calculated using the native structure factors (F_{av}) between 2.5-1.1Å resolution. The Harker sections for $x=0.50$, $y=0.50$ and $z=0.5$ are shown in Figure 1.2.3-1. The Fe-Fe vector is the largest feature in each section although there are other peaks of similar magnitude in each section. The S-S self vectors

and the Fe-S cross vectors are not present above the noise level of the map.

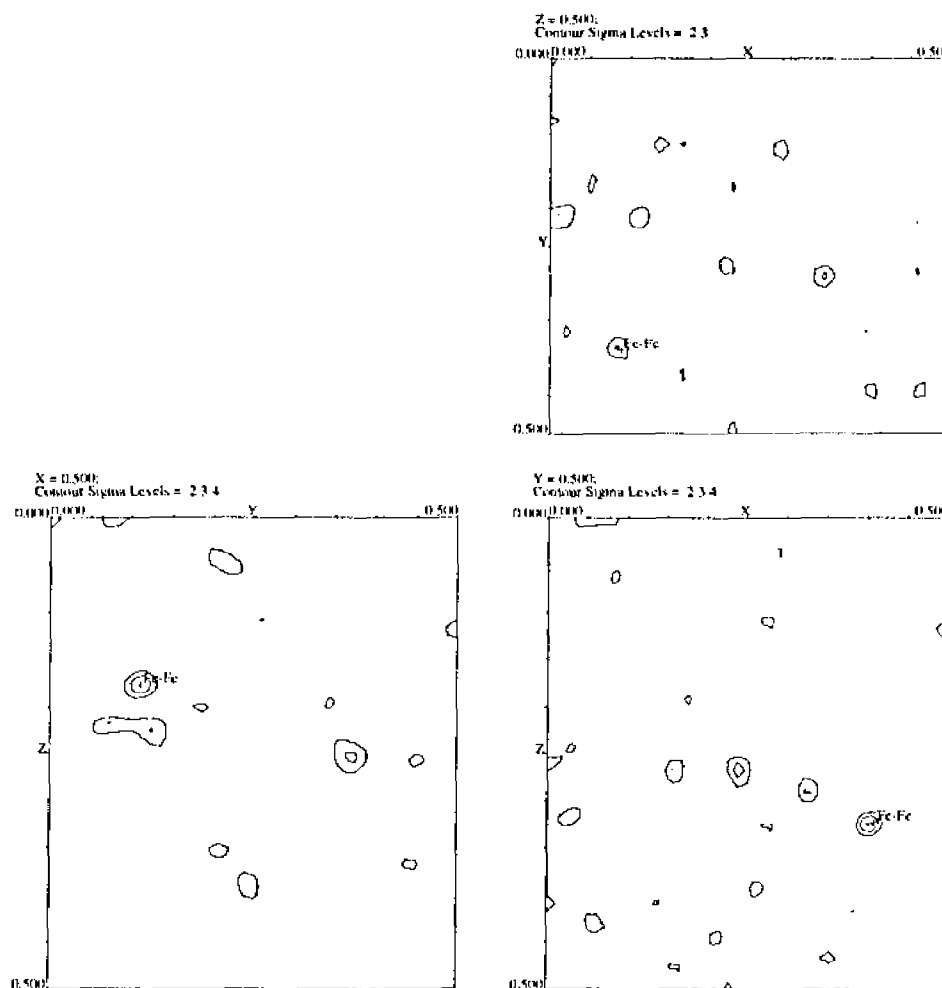


Figure 1.2.3-1. The Harker sections of the Patterson map calculated from the high resolution data using all data between 2.5Å and 1.1Å resolution. Only the asymmetric unit is shown. The peaks corresponding to Fe-Fe vectors are apparent and the predicted position of the Fe-Fe in each section vector is labeled.

Another Patterson map was calculated using the anomalous differences in the structure factors. Reflections between 4.0-1.0Å resolution were used with a filter applied to these reflections that discarded any reflections where the anomalous difference was greater than 40% of the structure factor for the *hkl* reflection. The map is shown in Figure 1.2.3-2. The Fe-Fe vectors are definitely the strongest features of this map and appear at the six sigma level in each Harker section. Again, the S-S vectors are not readily apparent above the two sigma level.

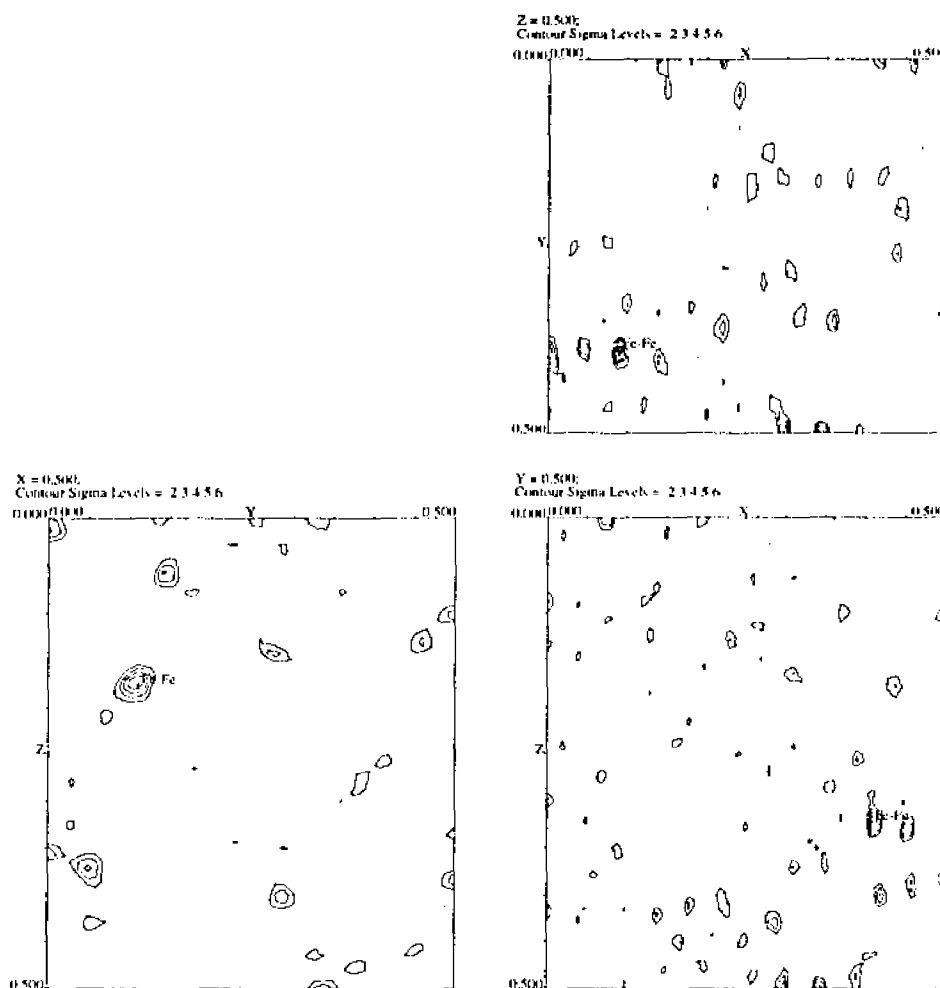


Figure 1.2.3-2. The Harker sections of the Patterson map calculated from the anomalous differences for all data between 4-1.0Å resolution. Only the asymmetric unit is shown. The peaks arising from Fe-Fe vectors are now quite obvious. The predicted position of the Fe-Fe vector is shown for each section.

Protein Phasing for anomalous data

The iron position was then used as though it was a heavy atom derivative atomic position, and the anomalous data, as though it was derivative data. The iron was input at the position derived from the Patterson map and protein phases were calculated with the Xtalview suite of programs. A map using the observed structure factors was generated from these phases and is shown in Figure 1.2.3-3. The four sulfur atoms appear in these maps at the 4σ level.

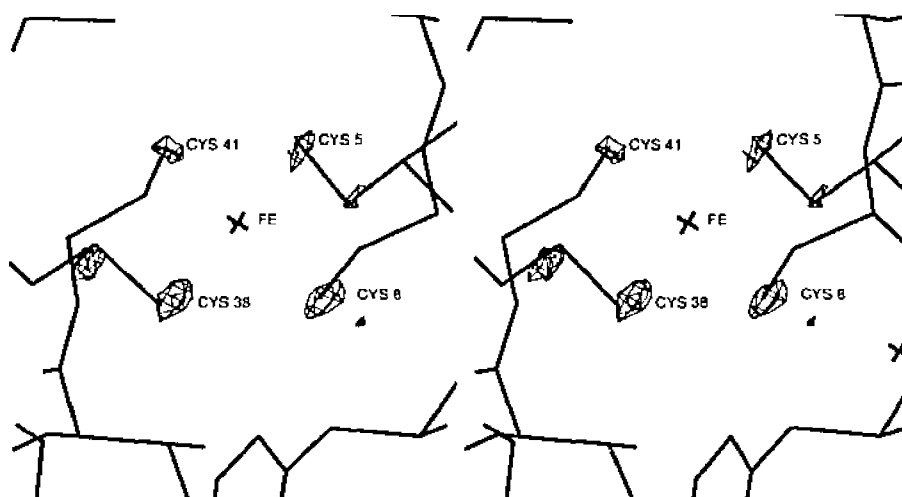


Figure 1.2.3-3. F observed map calculated from phases derived from the iron position.

References

- Crowther, R. A. (1972). A Method of Positioning a Known Molecule in an Unknown Crystal Structure. In *The Molecular Replacement Method*. (M. G. Rossman, eds.) pp. 173-178. Gordon and Breach, New York.
- Tronrud, D. E., Ten Eyck, L. F. and Matthews, B. W. (1987). "An Efficient General-Purpose Least-Squares Refinement Program for Macromolecular Structures." *Acta Crystallographica A* **23**, 489-501.

1.2.4 Low Resolution Refinement

Oxidized Form

The 337 atoms of the starting model (excluding the Fe atom) were refined as a rigid body against data in the resolution range between 11 and 2.5 Å, using the refinement package TNT. The initial model rotated by less than 0.1° and translated by less than 0.002 Å along any axis, lowering the R-factor from 0.45 to 0.41. Calculation of $F_O - F_C$ and $2F_O - F_C$ maps revealed the positions of the Fe atom and 9 of the 23 residue side chains not in common with the RdCp structure. The position indicated for the Fe atom was consistent with the strongest peaks observed in the native anomalous difference Patterson map (not shown). Refinement of the new model using the geometric restraints in TNT and a single overall temperature factor yielded $R=0.33$. Electron density difference maps ($F_O - F_C$ and $2F_O - F_C$) revealed the positions of six more residue side chains. At this point, manual adjustments were made to the existing structure using the computer graphics program TOM (Jones 1978). The new model was refined with TNT incorporating the 2243 reflections between 20-2.2 Å resolution with a resulting $R=0.27$. A subsequent set of electron density difference maps ($F_O - F_C$ and $2F_O - F_C$) established the positions of the remaining side chains. As before, manual adjustments were made to the model as needed. Alternating refinement of positions and temperature factors for the complete model including data between 20-2.2 Å resulted in an R-factor of 0.25.

Following the stereochemically restrained least squares refinement of RdPf with TNT, the slow-cooling simulated annealing protocol of X-PLOR (Brünger, Kuriyan et al. 1987; Brünger, Krukowski et al. 1990) was used for further refinement of atomic positions. Unit weights were applied to the structure factors during refinement. From a starting temperature of 4000 K, the final temperature of 300 K was reached by decreasing the temperature in steps of 25 K every 50 femtoseconds resulting in a residual of 0.26. Refinement of temperature factors in X-PLOR, followed by another round of simulated

annealing and temperature factor refinement, lowered the R-factor to 0.20 with rms. deviations from ideality in bond distances of 0.018Å and in bond angles of 3.25°.

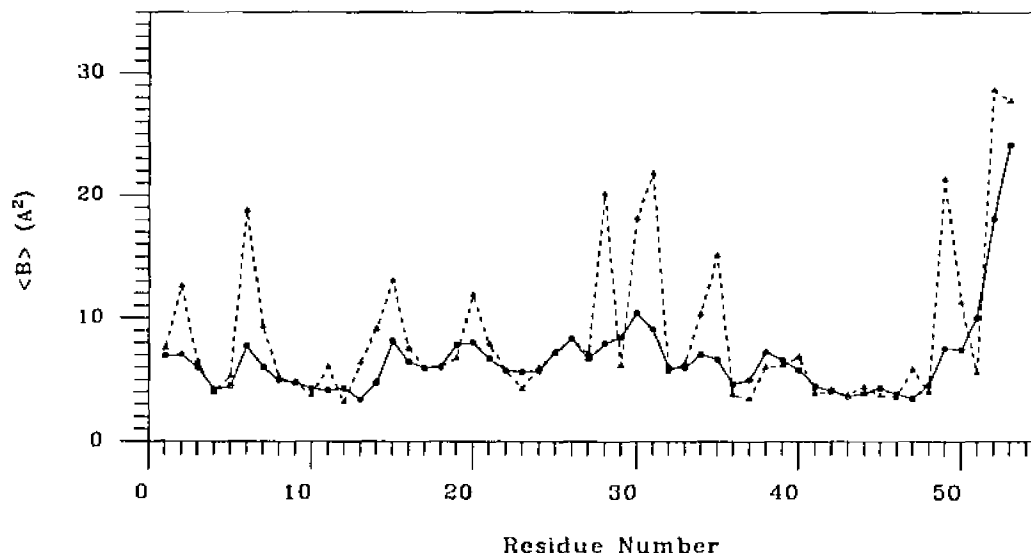


Figure 1.2.4-1. Average main chain and side chain temperature factors for the oxidized form. The main chain is shown by the solid line and the side chain is shown by the dashed line.

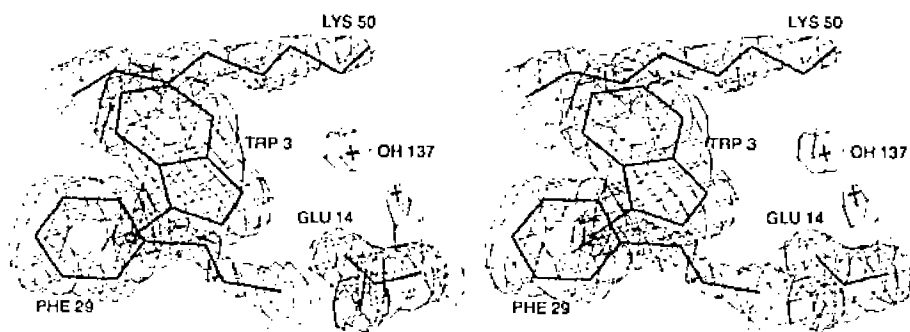


Figure 1.2.4-2. Stereo view of the final $2F_o - F_c$ electron-density map in the region around Glu 14 of the oxidized form. The map is contoured at the 2σ level.

The model was further refined against the data collected at -161°C , to a maximum resolution of 1.8Å (4439 reflections, 87% complete). The present refinement included all reflections with intensities greater than zero, collected with the detector centered at $2\theta=20^\circ$ (corresponding to a limiting resolution of 1.8Å). Electron density maps calculated to higher

resolutions (between 1.8Å and 1.1Å resolution) displayed pronounced anisotropic electron density due to smaller sampling of reciprocal space along the b^* direction. At present, the structure refinement, which includes the incorporation of 61 solvent H₂O molecules (of approximately 120 possible, based on volume considerations), has resulted in a residual of $R=0.178$, with rms. deviations from ideal bond lengths of 0.014Å and bond angles of 2.06°. The average temperature factors for the main chain and the side chain atoms are 6.6Å² and 8.9Å² respectively. Residues Glu-52 and Asp-53 at the C-terminus, and the side chains of Lys-6, Lys-28, Glu-30, Glu-31, and Glu-49, have large average B's (>17Å²) (see Figure 1.2.4-1). Two conformations of the Glu-52 and Asp-53 side chains were identified, although only the major conformation for each residue was refined. The final $2F_O-F_C$ electron density map in the region around Glu-14 is illustrated in Figure 1.2.4-2.

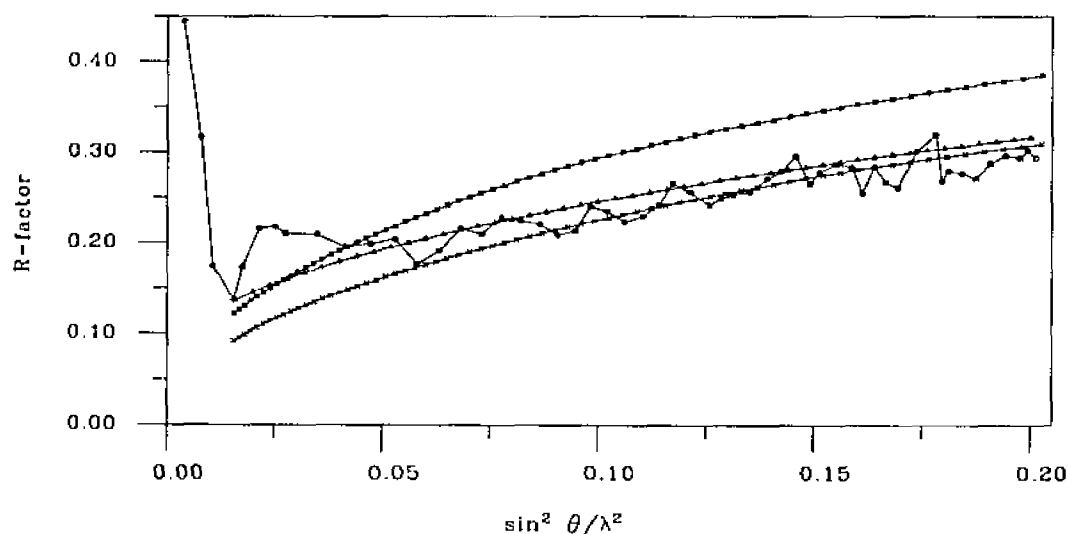


Figure 1.2.4-3. Dependence of R-factor on resolution for the oxidized form of rubredoxin. Theoretical curves are calculated for coordinate error of 0.15Å (top line) and 0.20Å (bottom line) and for a partial model containing 97% of the scattering mass with an average coordinate error of 0.15Å (middle curve). Only reflections in the resolution range 5 – 1.8Å were used in the refinement.

Estimation of the average coordinate error from a Luzzati (Luzzati 1952) plot was complicated by the nearly constant value of the R-factor as a function of resolution (Figure

1.2.4-3). This is possibly a consequence of applying unit weights to the structure factors during refinement (Adman 1990). If the dependence of R on resolution is examined to a limiting resolution of 1.1\AA where data between 1.8\AA and 1.1\AA resolution were not included in the refinement, then a Luzzati-type analysis (Srinivasan and Parthasarathy 1976) is consistent with an average coordinate error of $\sim 0.15\text{\AA}$, and a model that is $\sim 97\%$ complete. The missing scattering matter presumably consists of partially occupied solvent molecules that have not yet been included in the refinement.

A Ramachandran plot of the (ϕ, ψ) main chain torsion angles is shown in Figure 1.2.4-4. All residues have ϕ, ψ angles in allowed regions. The five residues with ϕ angles near 90° are glycines.

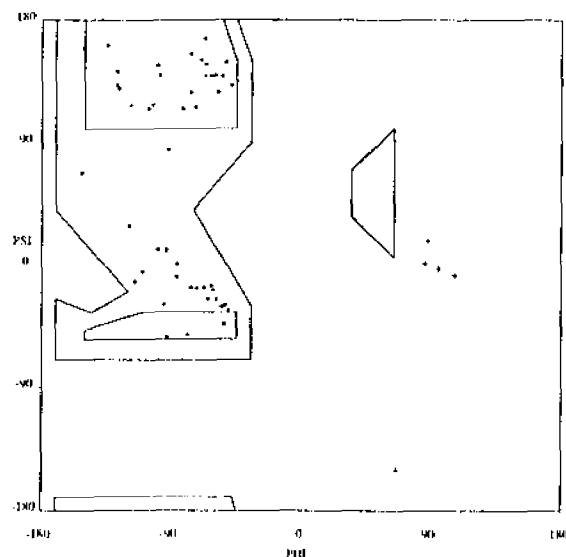


Figure 1.2.4-4. Ramachandran plot for the oxidized form of RdPf. Glycine residues are indicated by +.

Reduced Form

The R-factor between the data from the reduced form, and the oxidized RdPf observed data to 1.8\AA is 0.43. The corresponding R-factor with the RdPf oxidized model is 0.45. Difference Fourier maps between the data from the reduced crystal and, either the oxidized F_o 's or the oxidized model F_c 's, were very noisy and revealed no structural differences.

The model from the oxidized form, including solvent H₂O oxygen atoms, was then refined against the data from the reduced form as a rigid body resulting in $R=0.31$. The model rotated by 2.98° and translated by up to 0.31\AA along the crystallographic axes. The stereochemically restrained individual atomic positions were refined in the next cycles, followed by refinement of the temperature factors. The resulting model with $R=0.24$ was used to calculate F_O-F_C and $2F_O-F_C$ maps to evaluate the model. Again, no gross structural changes were observed, although it became apparent that the solvent region would require redetermination. A total of 24 solvent H₂O, all having either large temperature factors or questionable locations, were removed. The protein without the solvent shell was refined by simulated annealing with X-PLOR, followed by positional refinement of both the solvent shell and protein. At present, the R-factor is 0.193, with rms. deviations from ideal bond distances and bond angles of 0.012\AA and 1.95° , respectively. Electron density from the $2F_O - F_C$ map for this model is shown in Figure 1.2.4-5. The reduced form of RdPf has average temperature factors for the main chain atoms of $B=7.6\text{\AA}^2$ and $B=9.9\text{\AA}^2$ for side chain atoms. The side chains of the reduced RdPf with large average temperature factors ($B>15\text{\AA}^2$) include those mentioned above for the oxidized form, and Asp-20, Glu-31, Asp-34 and Asp-35 ($B>20\text{\AA}^2$) (see Figure 1.2.4-6). In contrast to the oxidized model, the side chains of residues 52 and 53 appear to have one major conformation.

A Ramachandran plot of the (Φ, Ψ) main chain torsion angles is shown in Figure 1.2.4-7. All 53 residues of the model have Φ, Ψ angles in the allowed regions. Again, the five residues with Φ angles near 90° are glycines.

Estimation of the average coordinate error by examination of the dependence of the residual on resolution is subject to the same complications discussed above for the oxidized model, namely the use of unit weights during refinement and the incomplete modeling of the solvent region. If the same type of analysis as used for the oxidized model is applied in

this case the estimated average coordinate error is approximately the same, i.e. 0.15\AA (see Figure 1.2.4-8).

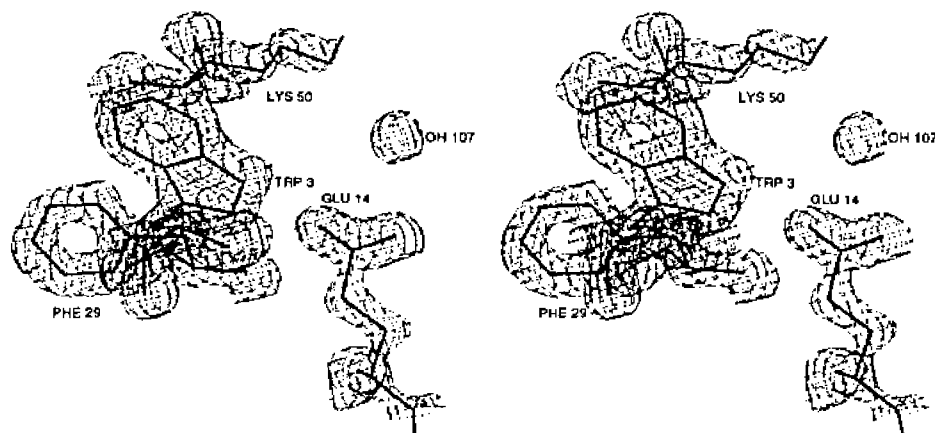


Figure 1.2.4-5. Stereo view of the $2F_o - F_c$ electron density map for the reduced form of RdPf in the area near Trp-3, Glu-14 and Phe-29. Contour level is 2σ .

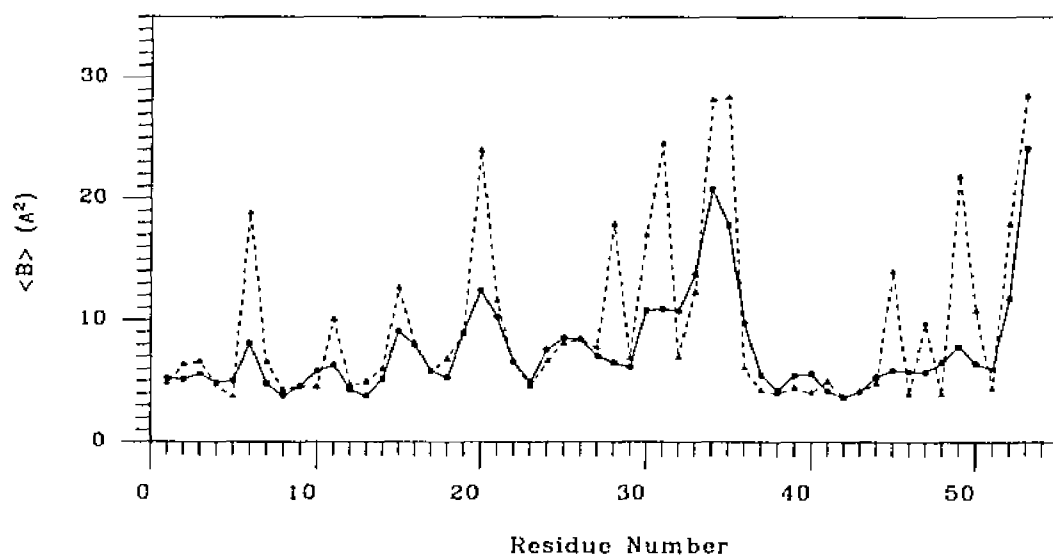


Figure 1.2.4-6. Average main chain and side chain temperature factors for the oxidized form. The main chain is shown by the solid line and the side chain is shown by the dashed line.

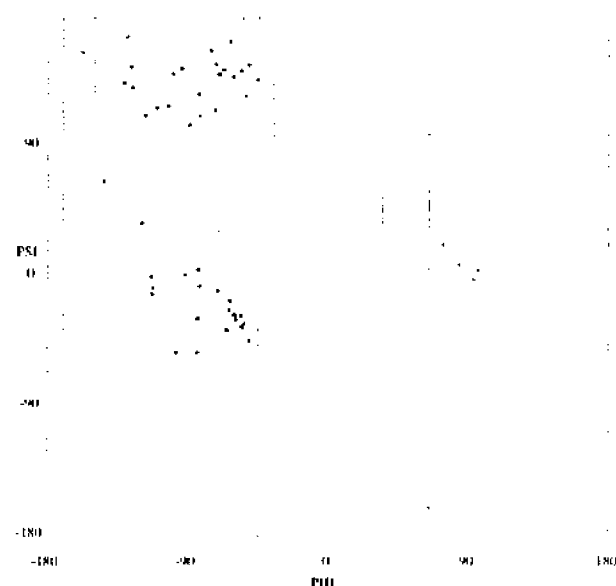


Figure 1.2.4-7. Ramachandran plot of the reduced form of RdPf. Residues represented by + are glycines.

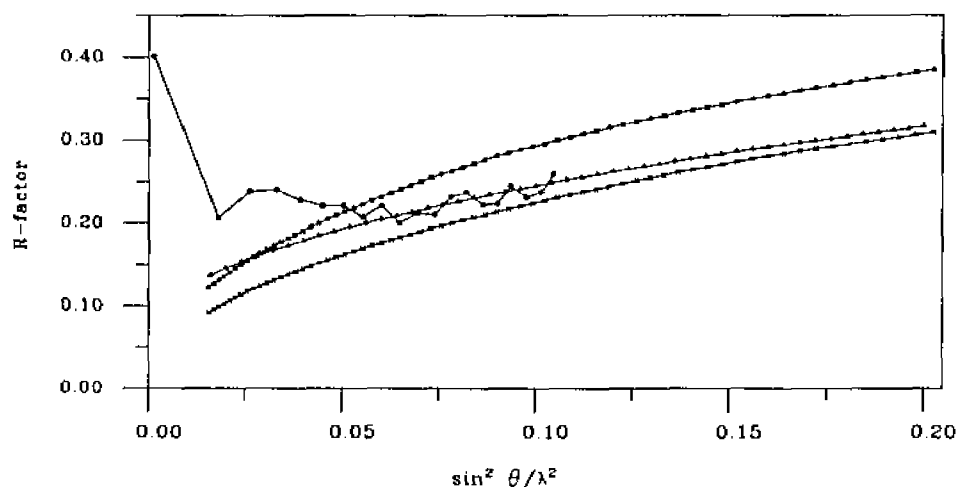


Figure 1.2.4-8. Dependence of R-factor on resolution for the reduced form of rubredoxin. Theoretical curves are calculated for coordinate error of 0.15 Å (top line) and 0.20 Å (bottom line) and for a partial model containing 97% of the scattering mass with an average coordinate error of 0.15 Å (middle curve). Only reflections in the resolution range 5 – 1.8 Å were used in the refinement.

References

- Adman, E. T. (1990). A tale of four iron-sulfur proteins: Sequence errors and other matters. Warrington, Science & Engineering Research Council, Daresbury Laboratory.
- Brünger, A. T., Krukowski, A. and Erickson, J. W. (1990). "Slow-cooling protocols for crystallographic refinement by simulated annealing." *Acta Crystallogr.* **46**, 585-593.

- Brünger, A. T., Kuriyan, J. and Karplus, M. (1987). "Crystallographic R Factor Refinement by Molecular Dynamics." *Science* **235**, 458-460.
- Jones, A. T. (1978). "A Graphics Model Building and Refinement System for Macromolecules." *Journal of Applied Crystallography* **11**, 268-272.
- Luzzati, V. (1952). "Traitement Statistique des Erreurs dans la Determination des Structures Cristallines." *Acta Crystallographica* **5**, 802-810.
- Srinivasan, R. and Parthasarathy, S. (1976). Some Statistical Applications in X-ray Crystallography. Oxford, Pergamon Press.

1.2.5 High Resolution Refinement

Oxidized Form

The coordinates from the PDB file for the oxidized rubredoxin (Bernstein, Koetzle et al. 1977; Abola, Bernstein et al. 1987) listed as ICAA were refined further against the data collected at -161°C to a limiting resolution of 1.1\AA (17125 reflections, 80% complete). The data used in this stage of the refinement was a combination of the data collected with the detector center set at $2\theta=20^{\circ}$ and $2\theta=60^{\circ}$ resulting in reflections with a maximum $2\theta=90^{\circ}$. However, since only 241 or approximately 3.5% of the possible reflections in the shell of reciprocal space between 1.1 and 1.0\AA resolution were successfully collected and integrated, these reflections were not included in the refinement.

It is interesting to note that significant portions of reciprocal space were not sampled within this 1.1\AA resolution shell. The indices in reciprocal space that were sampled were as follows; $0 \leq h \leq 30$, $0 \leq k \leq 24$ and $0 \leq l \leq 39$. This is not a result of anisotropic diffraction by the crystal, but is a result of data collection strategy and the limitations inherent in the geometry of the experiment. Considering that the **a** and **b** edges of the unit cell are very nearly the same length (33.8\AA and 34.6\AA respectively), it is apparent that the data is of a lower resolution (only 1.4\AA) along the *y*-axis than along the others. It is likely that the lack of high resolution data in reciprocal space along *k* is the cause of the anisotropic electron density seen in the maps. This hypothesis is consistent with the observation that the electron density is elongated along the *y*-axis (see Figure 1.2.5-1).

Two refinement packages were used in parallel for this stage of the refinement process. Both TNT (Tronrud, Ten Eyck et al. 1987) and XPLOR (Brünger, Kuriyan et al. 1987; Brünger, Krukowski et al. 1990) were used during four rounds of refinement and model building during which the refined model was evaluated against electron density maps calculated using the phases resulting from the refinements. A weighting scheme applying

$\frac{1}{\sigma^2(F_o)}$ weights to reflections between 10-1.1Å was used. The resulting model contained 104 solvent oxygens.

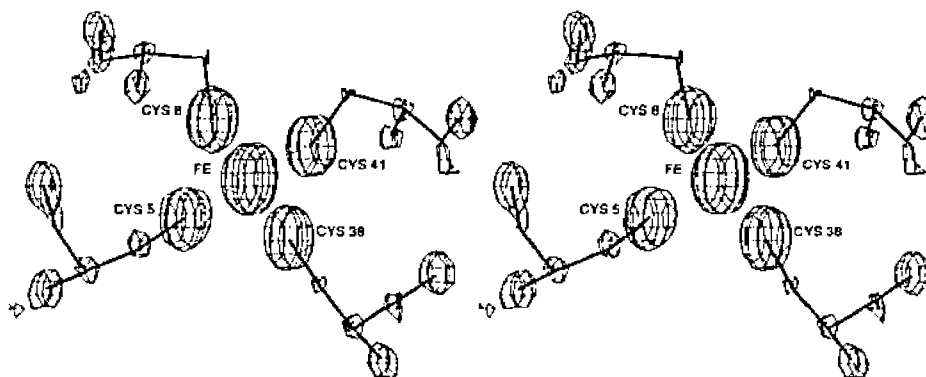


Figure 1.2.5-1. Stereo view at the 5 σ level of 2Fo-Fc electron density for the Fe-S center in rubredoxin showing the anisotropy of the density. The y-axis is vertical in the plane of the page.

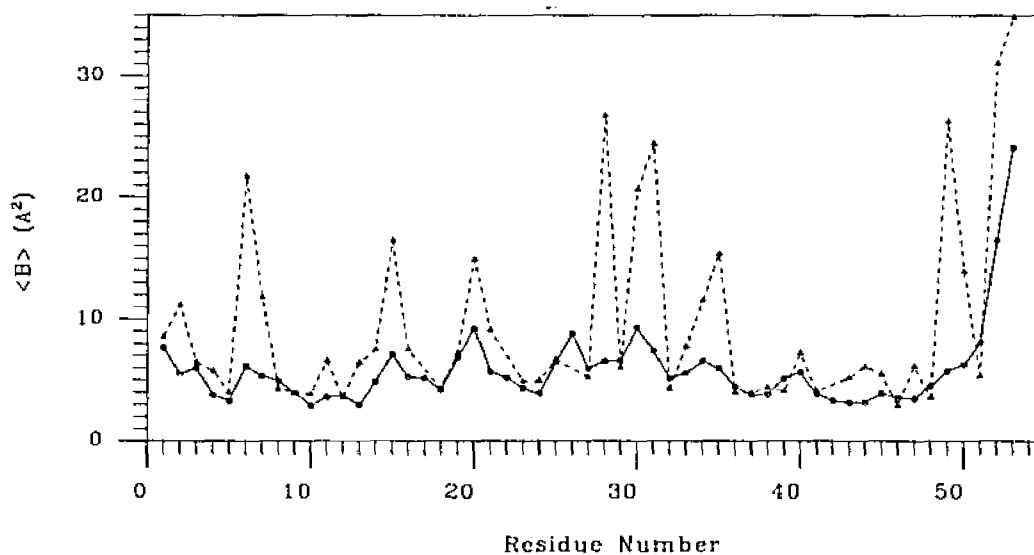


Figure 1.2.5-2. Average main chain and side chain temperature factors for the model of oxidized rubredoxin after refinement in XPLOR at between 10-1.1Å. The main chain is represented by the solid lines and the side chains by the dashed lines.

The XPLOR refined model resulted in a residual of R=18.0% with rms. deviations from target values for bond distance and bond angles of 0.014Å and 2.02° respectively.

The average temperature factors for the main chain and side chain atoms are 5.8\AA^2 and 9.7\AA^2 respectively while the average temperature factor for the solvent oxygens is 22.9\AA^2 . All solvent oxygen atoms were refined at full occupancy. Consistent with the low-resolution refinement results, the temperature factors for the side chain atoms of residues Lys-6, Lys-28, Glu-30, Glu-31 and Glu-49 have large B's ($>20\text{\AA}^2$) as do the two C-terminal residue (Glu-52 and Asp-53) as shown in Figure 1.2.5-2. No attempt was made to refine alternate conformations for these residues. Electron density calculated from the $2F_o - F_c$ map is shown in Figure 1.2.5-3 for the residues around Glu-14 (compare to Figure 1.2.4-2). Analysis of coordinate error using the method of Luzzati (Luzzati 1952) estimates the coordinate error to be approximately 0.14\AA on average (see Figure 1.2.5-4). Unlike before, in the low-resolution refinement, the analysis does not appear to be complicated by a nearly flat value of the R-factor as a function of the resolution. However, a weighting scheme was used as described above and this may explain a more normal distribution of R-factor vs. resolution. Additionally, the model is more complete than before (due to the addition of 40 solvent oxygens), and this should have a similar effect on the Luzzati plot.

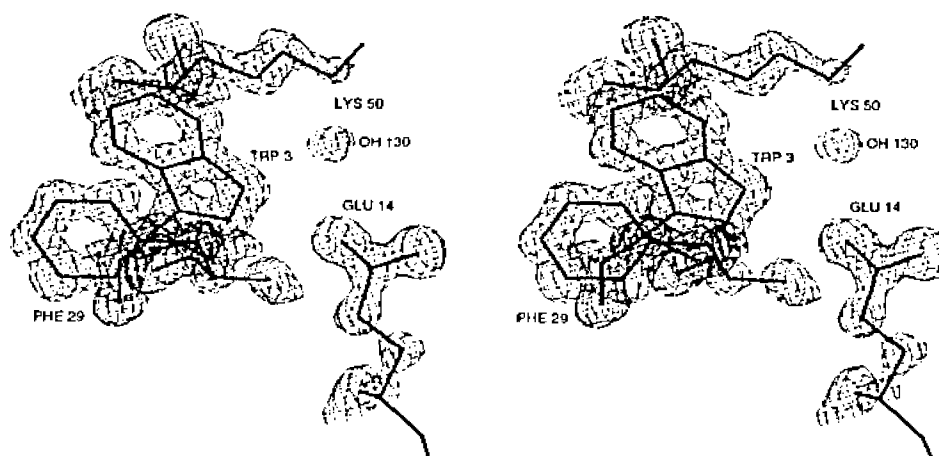


Figure 1.2.5-3. Stereo view of the refined $2F_o - F_c$ electron density map after refinement in XPLOR at 1.1\AA resolution. The region around Glu-14 and Trp-3 is shown at the two sigma level.

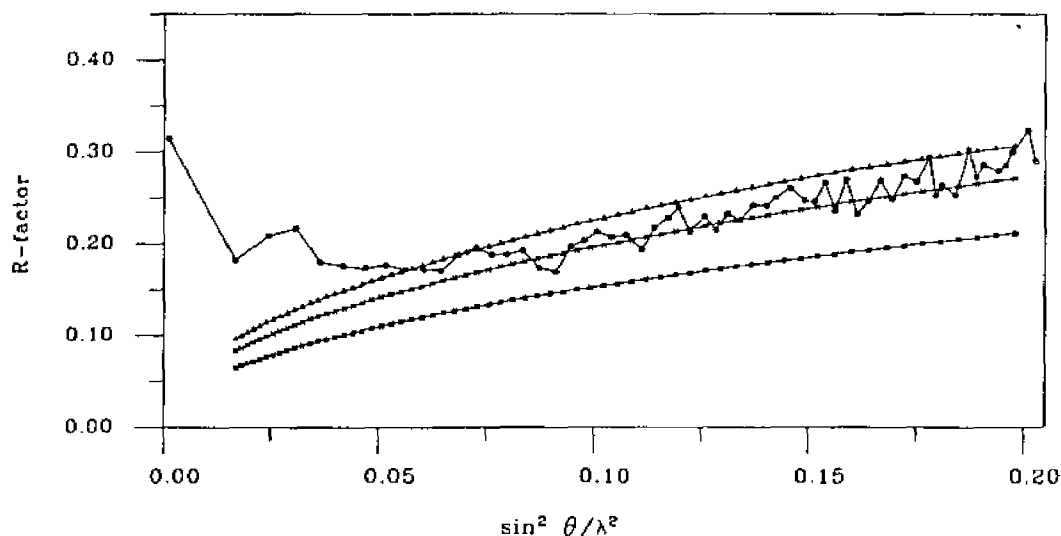


Figure 1.2.5-4. R-factor as a function of resolution after refinement in XPLOR. The theoretical curves are calculated for coordinate error of 0.15Å (top line), 0.13Å (middle line) and 0.10Å (bottom line). Reflections in the resolution range 8-1.1Å were used in the refinement.

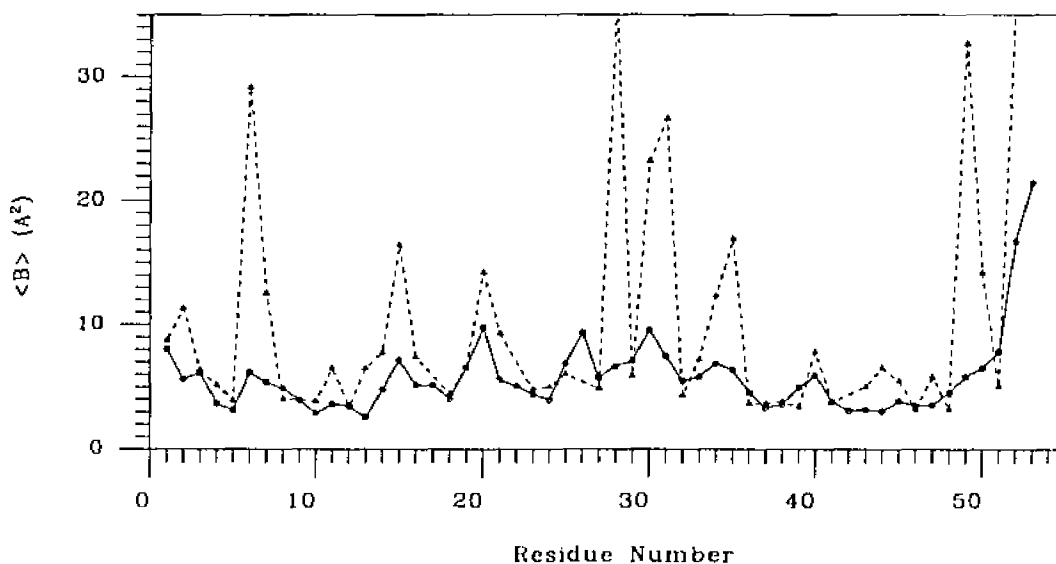


Figure 1.2.5-5. Average main chain and side chain temperature factors for the model of oxidized rubredoxin after refinement in TNT at between 10-1.1Å. The main chain is represented by the solid lines, and the side chains, by the dashed.

The refinement with TNT resulted in a model with a crystallographic residual of $R=16.1\%$ and a weighted residual of $R=13.1\%$. The rms. deviations of the bond distance

and bond angles from the target values are 0.017\AA and 2.98° respectively. Main chain atoms, side chain atoms and solvent oxygen temperature factors have average values of 5.8\AA^2 , 10.6\AA^2 and 22.8\AA^2 respectively (see Figure 1.2.5-5). Perhaps the most striking difference between refinement in TNT and XPLOR is reflected in the average values of the side chain B's for Lys-6, Lys-28, Glu-30, Glu-31, Glu-49, Glu-52 and Asp-53. The average values are higher in the model from TNT and results from one or two atoms of these side chains having much larger temperature factors than other atoms of the same residue. This appears to be a tendency of TNT, to be less effective than XPLOR at restraining the temperature factors of covalently bonded atoms to similar values. No attempt was made to model alternate side chain conformations. Electron density maps calculated from this model are virtually identical to those from the XPLOR refinement and are not shown. A slight complication occurs in the analysis or coordinated error using a Luzzati due to the choice of which residual to consider, weighted or unweighted (see Figure 1.2.5-6). The unweighted residual would seem to indicate an error that is essentially equal to that of the XPLOR refinement, i.e. 0.13\AA ; however, the R-factor curve from the weighted residuals is consistent with a coordinate error of 0.10\AA . It is doubtful that the coordinates derived from refinement in TNT are significantly more accurate than those resulting from the XPLOR refinement.

Reduced Form

The PDB coordinates for reduced RdPf refined at 1.8\AA resolution, listed as 1CAD, were refined against the entire data set of 7,859 reflections to a limiting resolution of 1.53\AA (91% complete). This is the same data set used in the low resolution (1.8\AA) refinement and the model that was derived from that refinement. The only difference was the inclusion of data between 1.8\AA and 1.5\AA resolution. This data is 98% complete to 1.6\AA and

approximately 55% complete in the shell between 1.5Å and 1.6Å. Reciprocal space was evenly sampled and includes $0 \leq h \leq 21$, $0 \leq k \leq 22$ and $0 \leq l \leq 27$.

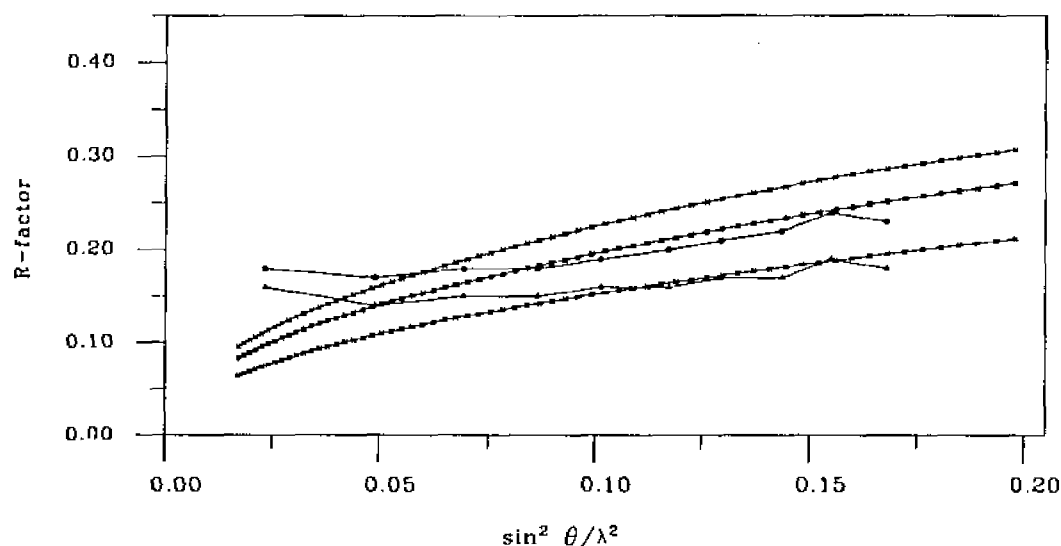


Figure 1.2.5-6. Plot of R-factor as a function of resolution from refinement with TNT. Two R-factors curves are shown, the upper curve results from units weights and the lower curve is the weighted R-factor. The theoretical curves are calculated for coordinate error of 0.15Å (top line), 0.13Å (middle line) and 0.10Å (bottom line). Reflections in the resolution range 8-1.1Å were used in the refinement.

As with the oxidized model a parallel refinement using both TNT and XPLOR was undertaken. However, only TNT was used in the final round of refinement. A weighting scheme using $\frac{1}{\sigma^2(F_o)}$ was applied during the refinement. The occupancy of the 110 solvent oxygens was refined during the final refinement round.

The resulting model containing a total of 524 non-hydrogen atoms refined to an unweighted crystallographic residual of $R=17.3\%$ and a weighted residual of $R_w=13.1\%$. Representative electron density from the final 2Fo-Fc map is shown for the residues surrounding Trp-3, and Glu-14 is shown in Figure 1.2.5-7 (compare to Figures 1.2.4-5 and 1.2.5-3). The geometry of the model deviates from the ideal values with rms. values of 0.022Å for bond distances and 3.78° for bond angles. The average values of the temperature factors for the main chain, side chain and solvent oxygen atom are 7.2Å²,

13.6Å² and 30.2Å² respectively (see Figure 1.2.5-8). A Luzzati analysis of the coordinate error estimates and average coordinate error of 0.13Å (see Figure 1.2.5-9).

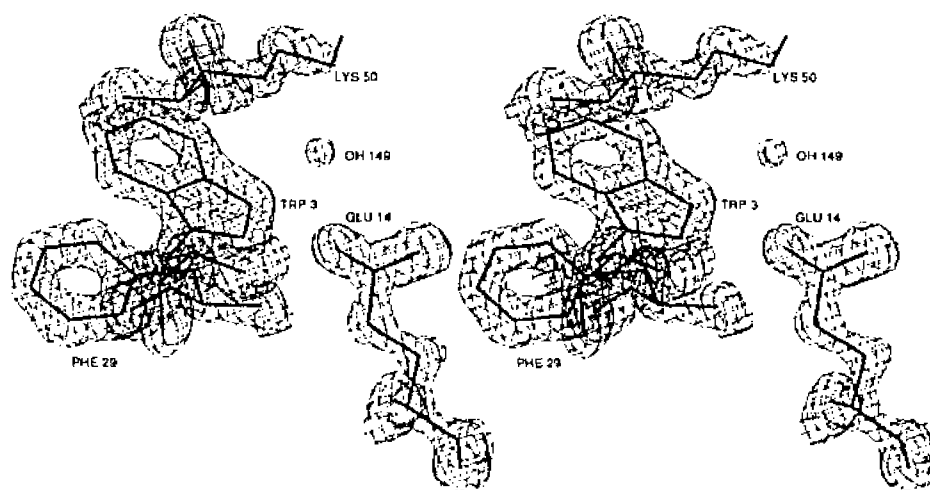


Figure 1.2.5-7. Stereo view of representative electron density in the 2Fo-Fc map from the final model in the area around Trp-3, Glu-14 and Phe-29. The contour level is 3 σ .

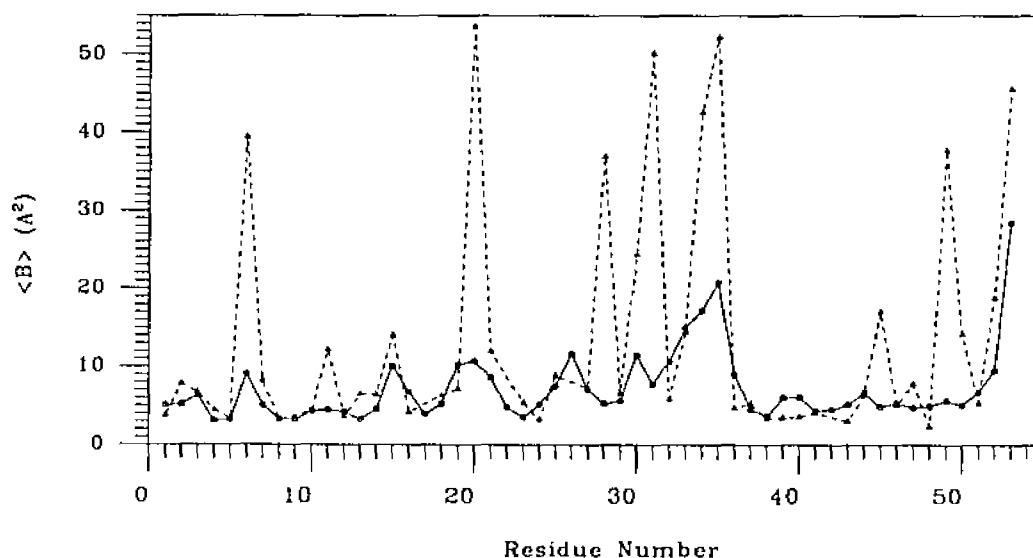


Figure 1.2.5-8. Average temperature factors for the model of reduced rubredoxin refined in TNT. Main chain atoms are shown in solid line and the side chain atoms are represented by the dotted line.

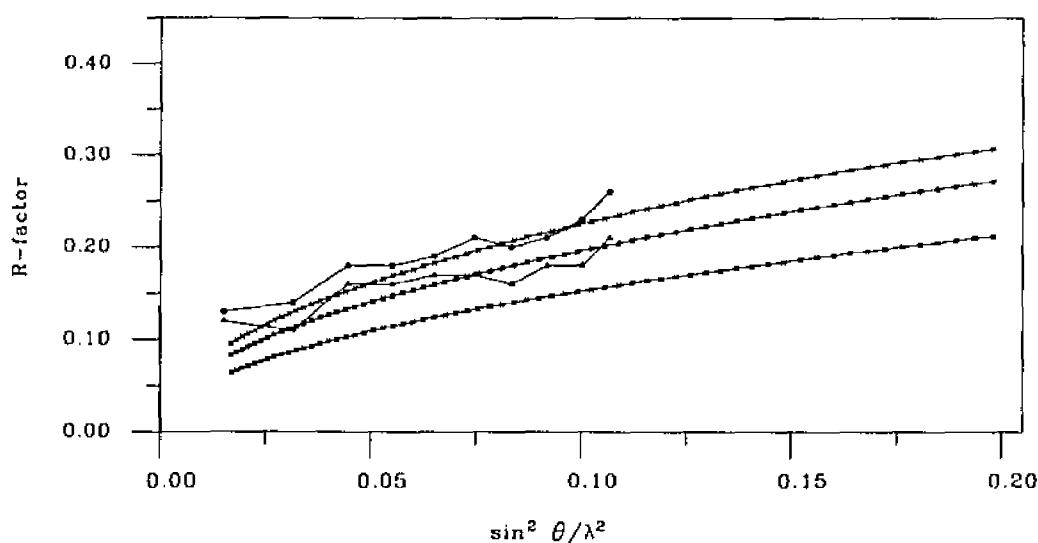


Figure 1.2.5-9. Plot of R-factor as a function of resolution from refinement of reduced RdPf with TNT. Two R-factors curves are shown, the upper curve results from units weights and the lower curve is the weighted R-factor. The theoretical curves are calculated for coordinate error of 0.15Å (top line), 0.13Å (middle line) and 0.10Å (bottom line). Reflections in the resolution range 8-1.5Å were used in the refinement.

References

- Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzle, T. F., et al. (1987). Protein Data Bank. Bonn/Cambridge/Chester, International Union of Crystallography.
- Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., et al. (1977). "The Protein Data Bank: A computer-based archival file for macromolecular structures." *J. Mol. Biol.* **112**, 535-542.
- Brünger, A. T., Krukowski, A. and Erickson, J. W. (1990). "Slow-cooling protocols for crystallographic refinement by simulated annealing." *Acta Crystallogr.* **46**, 585-593.
- Brünger, A. T., Kuriyan, J. and Karplus, M. (1987). "Crystallographic R Factor Refinement by Molecular Dynamics." *Science* **235**, 458-460.
- Luzzati, V. (1952). "Traitement Statistique des Erreurs dans la Determination des Structures Cristallines." *Acta Crystallographica* **5**, 802-810.
- Tronrud, D. E., Ten Eyck, L. F. and Matthews, B. W. (1987). "An Efficient General-Purpose Least-Squares Refinement Program for Macromolecular Structures." *Acta Crystallographica A* **23**, 489-501.

1.2.6 High Resolution Refinement with SHELXL-93

SHELXL-93 (Sheldrick 1993) is a program designed for the refinement of molecular structures against x-ray diffraction data. It has primarily been used for the refinement of small inorganic and organic molecules. However, the ease with which restraints can be applied to the refinement process makes the use of SHELXL-93 an attractive choice for the refinement of macromolecular (i.e., protein) structures as well. The application of restraints for the refinement of a protein is facilitated by a utility program (PDBINS) that accepts a standard PDB coordinate set and writes an appropriate SHELXL-93 input file (*.INS) containing the restraints. The restraints are those compiled by Engh and Huber (Engh and Huber 1991) from data in the Cambridge Data Base (Allen, Kennard et al. 1983). PDBINS also sets up default values for the correlation of temperature factors between atoms that are covalently bonded to each other and for the corrections applied to the atomic scattering factors required in order to account for the presence of bulk solvent in the crystal.

This program is attractive for the refinement of RdPf because it offers the possibility to refine the structure by the conjugate-gradient (CGLS) algorithm (Hendrickson and Konnert 1980) or by full-matrix least squares (LSFM) with or without restraints. Isotropic or anisotropic displacement parameters can be used in the refinement. Furthermore, hydrogen atoms can be included during the refinement if they are warranted. Additionally, a free R-factor refinement can be performed allowing an additional check against over-refining the structure. Considering that the rubredoxin model could be expected to contain approximately 540 non-hydrogen atoms (414 protein, ~120 solvents) per asymmetric unit, refinement with isotropic B's (4 parameters/atom) would result in a data to parameter ratio of approximately 8:1 (17800 data/2136 parameters, depending on how many reflections were removed for calculation of the free R-factor) for the oxidized form. Refinement of the protein with anisotropic displacement parameters (9

parameters/atom) while keeping isotropic B's for the solvent atoms would maintain a data to parameter ratio of 4 to 1 without restraints and 5 to 1 with restraints. Anisotropic refinement of the reduced form (7780 data) does not seem to be prudent although a restrained least-squares full-matrix refinement appears to be within reach. A particular advantage of refinement by the method of least squares would be the possibility of obtaining coordinate, bond distance and bond angle errors directly from the least-squares parameter matrix. It is of interest to compare these values with errors obtained in the usual way for protein structures, i.e., a Luzzati (Luzzati 1952) plot of R-factor vs. resolution and comparison with a standard dictionary.

Considering the multitude of possibilities available with SHELXL-93, a test was conducted to evaluate the results obtained from each possibility. The possibilities to be evaluated were CGLS or LSFM refinement, either with or without restraints, using isotropic or anisotropic B's and using one of two weighting schemes. The choice of weighting schemes was between the weights generally used in a crystallographic refinement, i.e., $w = \frac{1}{\sigma^2(F_o^2)}$ and between $w = \frac{1}{\sigma^2(F_o^2) + (aP)^2 + bP}$ where $P = \frac{F_o^2 + 2F_c^2}{3}$ and the constants **a** and **b** are chosen so as to give a flat analysis of the variance between the observed and calculated square of the structure factors over the full range of observations. The program suggests values of these constants as a result of its analysis of the variance after the refinement. For this test only two values of these constants were used, i.e., $a=b=0$ (which reduces to $w = \frac{1}{\sigma^2(F_o^2)}$) or $a=0.1$ and $b=0$ (the default values).

The test was applied to the oxidized form of RdPf without hydrogen atoms included. Refinement included all data within the 8.0Å resolution shell. This resolution cutoff excluded 73 low angle reflections. An additional 10% of the reflections (every 10th in the list, 1732 reflections) were removed from the data set and not used during the refinement in order to calculate a free R-factor. The starting coordinates were those resulting from the high resolution refinement with XPLOR. For each case 20 refinement

cycles were done, and the default restraints were used for the restrained refinements. The results were evaluated on the basis of R-factor, goodness of fit (GoF), R(free) and the errors in the bond distances, bond angles and atomic coordinates. For LSFM refinement the errors from the error matrix were also considered. Unrestrained refinement with anisotropic displacement parameters tended to be unstable and therefore was not included in the test. The results are summarized in Table 1.2.6-1.

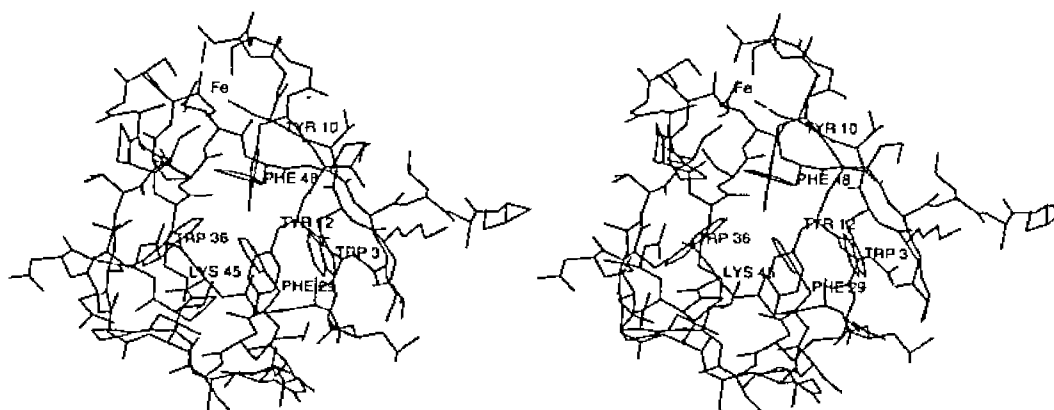


Figure 1.2.6-1. Stereo view of the rubredoxin after unrestrained refinement in SHELXL-93 showing that to a large degree the molecule is intact and visually adequate. The aromatic residues of the core are labeled and appear to be nearly planar. Many atoms in surface residues are not within bonding distance of each other. The solvent shell (not shown) is not greatly affected by unrestrained refinement. Two residues at the C-terminus (right side of figure) are visibly disrupted.

Unrestrained refinement did not appear to be a reasonable option for three reasons. Firstly, the initial intent was to use anisotropic displacement parameters in order to identify any regions of the molecule that might be undergoing large correlated motions and the simple observation that refinement was unstable under these conditions precluded the possibility of unrestrained refinement. Second, although the GoF tended to be better, (especially for $w = \frac{1}{\sigma^2(F_o^2)}$ weighting) the crystallographic residuals R(free) and R(all data) tended to be ~0.3-0.5% higher than in the restrained refinement. The third and most important reason arises from the fact that the refined model must be chemically reasonable and that the calculated errors be small. Although the model remained intact and recognizable to the eye (see Figure 1.2.6-1), the rms deviation in bond distances and angles

from chemically reasonable values is unacceptably high (see Table 1.2.6-1). For LSFM refinement the errors calculated from the error matrix were near double those in the restrained refinement.

Table 1.2.6-1. Results of the evaluation of refinement methods in SHELXL-93. The reflection list included 17366 reflections, 15561 of which were used in the refinement and 1732 which were withheld in order to calculate a free R-factor. A resolution range of 8-1.0Å was used which excluded 74 low angle reflections. For each situation 20 cycles of refinement were performed .

Isotropic displacement parameters				
Value	CG Least Squares		Full Matrix Least Squares	
	Restrained	Unrestrained	Restrained	Unrestrained
# of Parameters	2074	2073	2074	2073
# of Restraints	1656	—	1657	—
Weighting Scheme	default	default	default	default
Initial GoF	13.38	13.38	13.38	13.38
Initial GoF (restrained)	13.43	—	13.43	—
GoF	4.41	4.32	4.39	4.29
GoF (restrained)	4.19	—	4.18	—
R (refined data)	17.77%	17.58%	17.69%	17.43%
R (all data)	18.40%	18.89%	18.38%	18.81%
R (4 sigma, 14722 refs)	17.27%	17.07%	17.19%	16.91%
R (free)	20.37%	20.70%	20.68%	21.34%
rms Bond	0.016Å	0.113Å	0.015Å	0.247Å
average Bond sigma			0.021Å	0.041Å
rms Angle	1.91°	7.01°	1.87°	9.04°
average Angle sigma			1.66°	2.72°
Luzzati plot	~0.15Å	~0.15Å	~0.15Å	~0.15Å
average coordinate sigma			0.056Å	0.072Å
Weighting Scheme	1/ σ^2	1/ σ^2	1/ σ^2	1/ σ^2
Initial GOF	27.15	28.49	27.17	28.49
Initial GOF (restrained)	27.23	—	27.23	—
GOF	24.77	19.08	25.09	15.35
GOF (restrained)	23.53	—	23.83	—
R (refined data)	18.61%	18.43%	18.55%	18.26%
R (all data)	19.41%	19.79%	19.38%	19.70%
R (4 sigma, 14722 refs)	18.01%	17.73%	17.95%	17.55%
R (free)	22.39%	22.38%	22.61%	22.91%
rms Bond	0.020Å	0.347Å	0.021Å	0.217Å
average Bond sigma			0.021Å	0.043Å
rms Angle	2.21°	10.13°	2.21°	9.78°
average Angle sigma			1.89°	3.10°
Luzzati plot	~0.15Å	~0.15Å	~0.15Å	~0.15Å
average coordinate sigma			0.054Å	0.075Å

Table 1.2.6-1 (cont.)

Anisotropic displacement parameters				
Value	CG Least Squares		Full Matrix Least Squares	
	Restrained	Unrestrained	Restrained	Unrestrained
# of Parameters	4144		4144	
# of Restraints	4778		4778	
Weighting Scheme	default		default	
Initial GOF	14.55		14.55	
Initial GOF (restrained)	15.62		15.62	
GOF	3.96		3.93	
GOF (restrained)	3.41		3.39	
R1 (refined data)	14.88%		14.83%	
R1 (all data)	15.62%		15.59%	
R1 (4 sigma, 14722 refs)	14.47%		14.42%	
R1 (free)	18.35%		18.70%	
rms Bond	0.016Å		0.016Å	
average Bond sigma			0.021Å	
rms Angle	1.94°		1.93°	
average Angle sigma			1.55°	
Luzzati plot	~0.15Å		~0.15Å	
average coordinate sigma			0.050Å	
Weighting Scheme	1/ σ^2		1/ σ^2	
Initial GOF	30.40		30.40	
Initial GOF (restrained)	32.65		32.65	
GOF	16.41		16.25	
GOF (restrained)	14.07		13.92	
R1 (refined data)	15.88%		15.76%	
R1 (all data)	16.73%		16.63%	
R1 (4 sigma, 14722 refs)	15.34%		15.20%	
R1 (free)	20.44%		20.58%	
rms Bond	0.019Å		0.019Å	
average Bond sigma			0.021Å	
rms Angle	1.75°		2.57°	
average Angle sigma			1.74°	
Luzzati plot	~0.15Å		~0.15Å	
average coordinate sigma			0.051Å	

Analysis of the results with respect to weighting scheme suggested that it would be prudent to use the default weighting scheme where the weights would be adjusted based on the difference between F_o and F_c and to use the values for the constants a and b that were suggested by the program rather than adjust the values manually until the numbers were perfect. This approach makes sense on a couple points. First, the free R-factors (in fact all of the R-factors) were lower with the default weighting scheme. The second point is that the sigma's for each reflection are suspect to begin with since they are generally considered to be underestimated. In a sense it may be reasonable to increase the sigma by some amount in accordance with the agreement between F_o and F_c . Additionally, we know that low angle reflections are affected by the solvent disorder in the crystal so they could certainly be downweighted and this scheme will have the most effect on the strongest reflections. This may indeed be a numbers game because the geometry and errors are essentially the same one way or the other. Nonetheless, a low R-factor and a GoF near 1.0 are what most people seem to look for in a refinement and since the analysis of the structure is not going to be affected (at least in this case where restraints were also applied), the decision was made in to not use $\frac{1}{\sigma^2(F_o^2)}$ weights.

Visual inspection of the electron density maps did not reveal any significant differences between CGLS and LSFM refinement. Likewise, the choice of weights had no palpable effect on the maps. One feature of the maps is striking however; atomic positions are clearly visible as unconnected balls of electron density (see Figure 1.2.6-4) instead of connected tubes of density enclosing the bond between atoms, as is usually seen at the diffraction limit for proteins. It should be noted that this feature is not unique to refinement in SHELXL-93, but is also seen in maps calculated from coordinates refined in TNT and XPLOR. Although this is striking, it should not be regarded as unusual for structures refined to this resolution. Another feature that is worth noting and has bearing on the

refinement strategy is the evidence of hydrogen atoms in Fo-Fc maps at the 3σ level (see Figure 1.2.6-2).

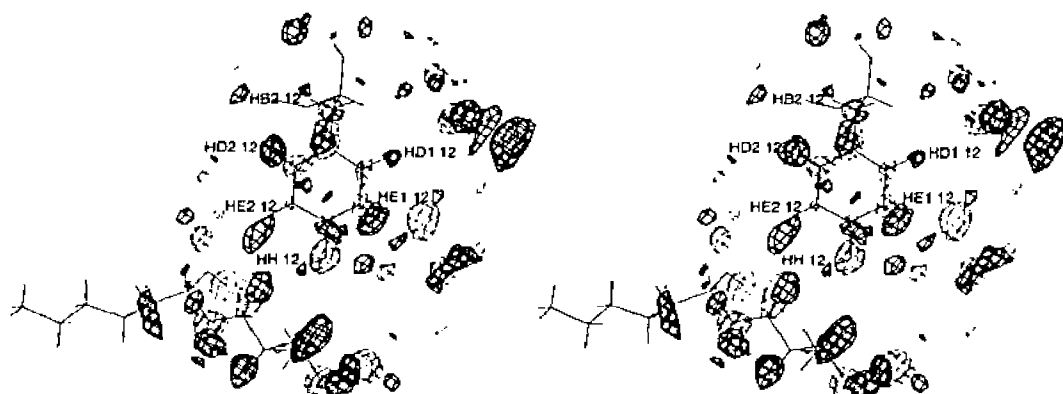


Figure 1.2.6-2. Stereo view of the electron density map around Tyr-12. The light lines are from the $2Fo-Fc$ map at 4σ , and the heavy lines are from the $Fo-Fc$ map at the 2σ level. Although noise is evident at the 2σ level in the $Fo-Fc$ map, there is clearly density in places where hydrogen atoms can reasonably be expected to be found.

Based on these observations the decision was made to refine both the oxidized and reduced form of RdPf using the restrained LSM method and the default weighting scheme. Restraints were not applied to the Fe-S interactions. Throughout the refinement process default values were used for all restraints. The values of the constants a and b in the weights were left at the default values until the final refinement cycles when all the solvent atoms had been located. At that time, the values suggested by the program were applied. In the case of the oxidized form, the data seemed to support the inclusion of hydrogen atoms subject to riding restraints. The refinement of the reduced form did not include hydrogen atoms, and all atoms were refined isotropically.

Oxidized Form

The model from XPLOR was subjected to five rounds of restrained full matrix least-squares refinement of 4256 parameters against 15561 reflections and 4783 restraints. There was no manual intervention during this refinement process except for inspection of atoms in the solvent shell. The difference Fourier map of the asymmetric unit was searched

and difference peaks that were an appropriate distance from existing atomic positions were automatically added as water. At the end of each refinement cycle, solvent oxygens with $B(\text{iso}) \geq 50.0 \text{ \AA}^2$ were removed from the coordinate list. The final model contains 414 non-hydrogen protein atoms (with anisotropic parameters) and 132 solvent oxygen atoms (with isotropic B's). Hydrogen atoms were included as riding atoms on the protein atoms only. The final refinement cycle resulted in a residual of $R=13.9\%$ (free $R=17.2\%$) and $\text{GoF}=1.08$ with an rms deviation of bond distances and angles from ideality of 0.014 \AA and 1.75° respectively. These values are in good agreement with the errors derived from the least squares parameter matrix where the average sigma for bond distances and angles are 0.021 \AA and 1.48° respectively. A plot of the average $B(\text{eq})$'s is shown in Figure 1.2.6-3. Large values (i.e., $B(\text{av}) \geq 20 \text{ \AA}^2$) are observed for the sidechains of residues Lys-6, Lys-28, Glu-31, Glu-52 and Asp-53.

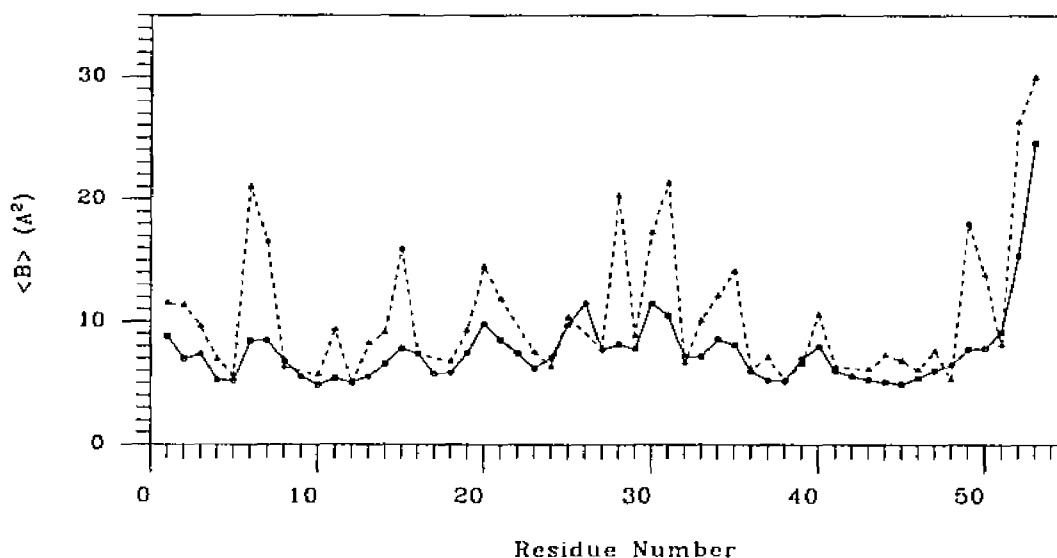


Figure 1.2.6-3. Plot of average equivalent B-factors for the oxidized form of RdPf after refinement in SHELXL-93 with anisotropic displacement parameters and hydrogen atoms included. Main chain atoms are represented by the solid line, and side chain atoms, by the dashed lines.

For comparison to previous refinements, electron density around Trp-3, Glu-14, Phe-29 and Lys-50 from the $2F_o - F_c$ map is shown in Figure 1.2.6-4. The density shown in this figure is contoured at the two sigma level, and it is apparent that the individual atoms

are starting to become defined. This density is representative of the average side chain density at this level. All of the main chain atoms, except two atoms of residue Asp-53, are in 3σ density, as are nearly all of the side chain atoms. Figure 1.2.6-5 shows electron density for four residues in the hydrophobic core, Tyr-10, Tyr-12, Trp-36 and Phe-48, from the 2Fo-Fc map at the 4σ level. The atoms are resolved as individual spheres, and the oxygen atoms of the tyrosines and the nitrogen atom of the indole ring appear larger than the carbon atoms.

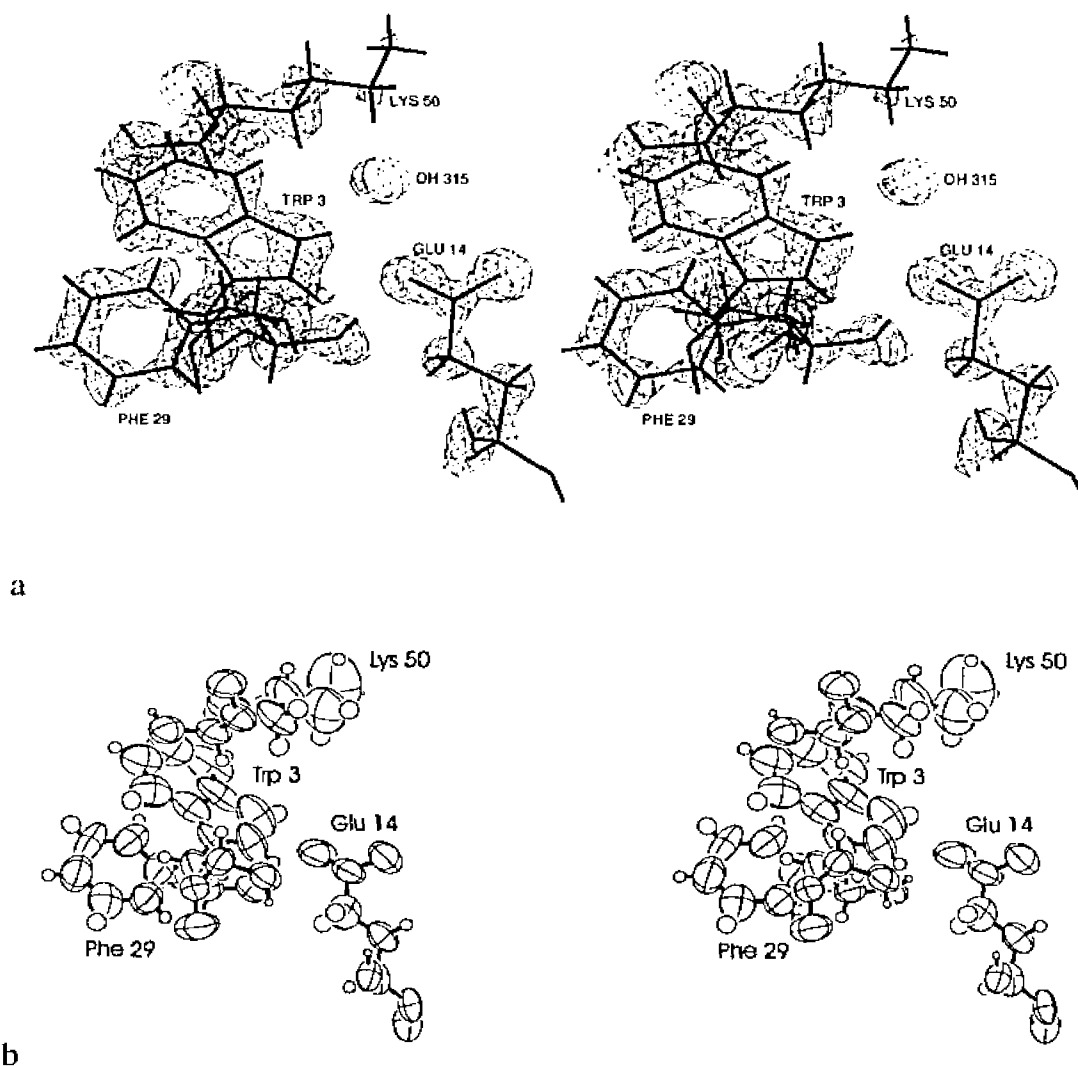


Figure 1.2.6-4. a) Electron density from the 2Fo-Fc map around Trp-3, Glu-14, Phe-29 and Lys-50 at the 2σ level. Compare with Figures 1.2.4- and with 1.2.5- , b) ORTEP representation of the same region showing the thermal ellipsoids at the 60% level.

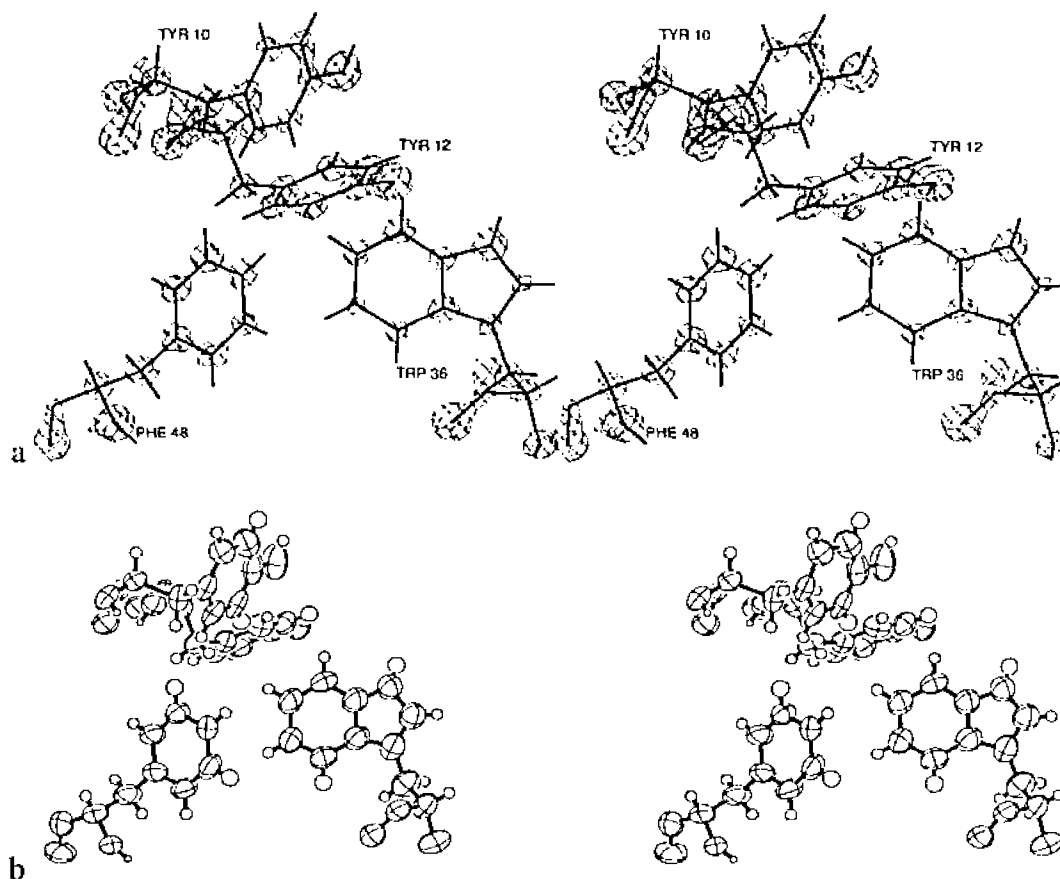


Figure 1.2.6-5. a) Electron density at 4σ from the $2F_o-F_c$ map for residues in the hydrophobic core. Many of the atoms are resolved as individual spheres and the oxygen and nitrogen atoms are significantly larger than the carbon atoms. b) ORTEP drawing of the hydrophobic core showing 60% atomic ellipsoids.

As a consequence of LSFM refinement, the average coordinate error for this model can be estimated in two ways. A plot of the R-factor versus the resolution as shown in Figure 1.2.6-6 would indicate the average coordinate error is approximately 0.06\AA . This is in very good agreement with the average error in the atomic coordinates of 0.047\AA as calculated from the least-squares parameter matrix. This result would seem to indicate that a Luzzati type analysis of coordinate error is accurate for a complete model and probably overestimates the coordinate error for an incomplete model.

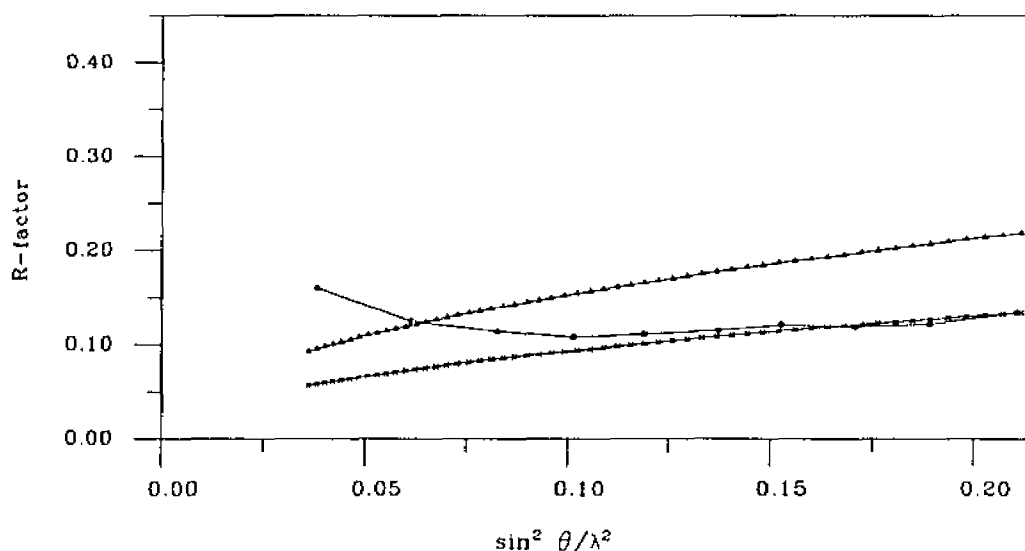


Figure 1.2.6-6. A plot of R-factor versus resolution for the SHELXL-93 refined model of RdPf. Theoretical curves are shown for an estimated error of 0.10 Å (top line) and 0.06 Å (bottom line).

Reduced Form

The model of reduced form TNT was taken through 4 rounds of restrained LSFM refinement against 7008 data and 1686 restraints. The 2138 parameters of the model were refined in the resolution range between 8-1.5 Å which resulted in 73 reflections with $d > 8.0$ Å being excluded. An additional 778 reflections were removed before refinement to facilitate the calculation of a free R-factor. The refinement was done in the same fashion as for the oxidized form (see above), with the exceptions that the hydrogen atoms were not included and that the temperature factors were refined isotropically. The final model contains 130 solvent oxygen atoms, in addition to the 414 non-hydrogen atoms of the protein. The resulting residual is $R=16.5\%$ against all 7859 data ($GoF=1.09$) and the free R-factor is 22.9% for the 778 reflections excluded from the refinement. The rms deviation from ideality of bond distances and angles is 0.013 Å and 1.81° respectively. Again these values are in good agreement with the errors derived from the LSFM error matrix. The average errors in the distances and angles were calculated to be 0.022 Å and 2.15°. The average B-factors of the reduced form are shown in Figure 1.2.6-7. The side chains of

Lys-6, Asp-20, Lys-28, Glu-30, Glu-31, Asp-34, Asp-35, Glu-49, Glu-53 and Asp-53 have average temperature factors $>20\text{\AA}^2$.

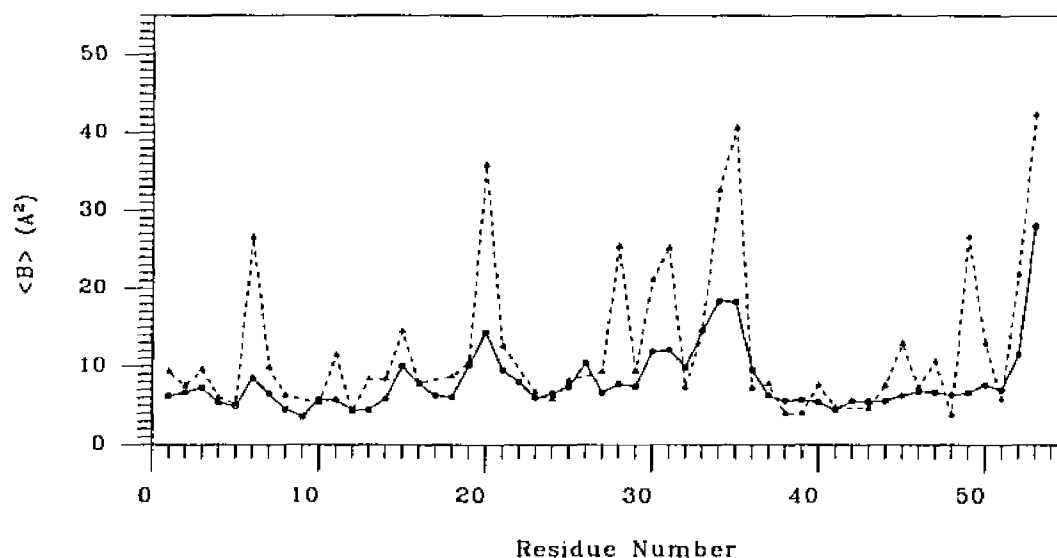


Figure 1.2.6-7. Plot of average temperature factors for the reduced form of RdPf after refinement in SHELXL-93 at 8-1.5 \AA resolution. The solid line represents the main chain atoms, and the dashed line represents the side chain atoms. For a comparison with the other refinements, see Figures 1.2.4-6 and 1.2.5-5.

A plot of the R-factor as a function of resolution is shown in Figure 1.2.6-8 as a measure of the average coordinate error. The plot shows three theoretical lines, one for a partial model (97% complete with $\sigma=0.08\text{\AA}$) and two for complete models with 0.10\AA error (bottom line) and 0.12\AA error (top line). The curves for the complete model would suggest an average error in the coordinates of between 0.10\AA and 0.12\AA . The average error derived from the least squares refinement matrix indicate the error to be 0.08\AA , which fits the middle line well if the model is only 97% complete. This apparent discrepancy may arise from unmodeled disorder in both the solvent shell and the protein side chains. Electron density from the 2Fo-Fc map in the area around two of the unique residues of RdPf, Trp-3 and Glu-14 is shown in Figure 1.2.6-9.

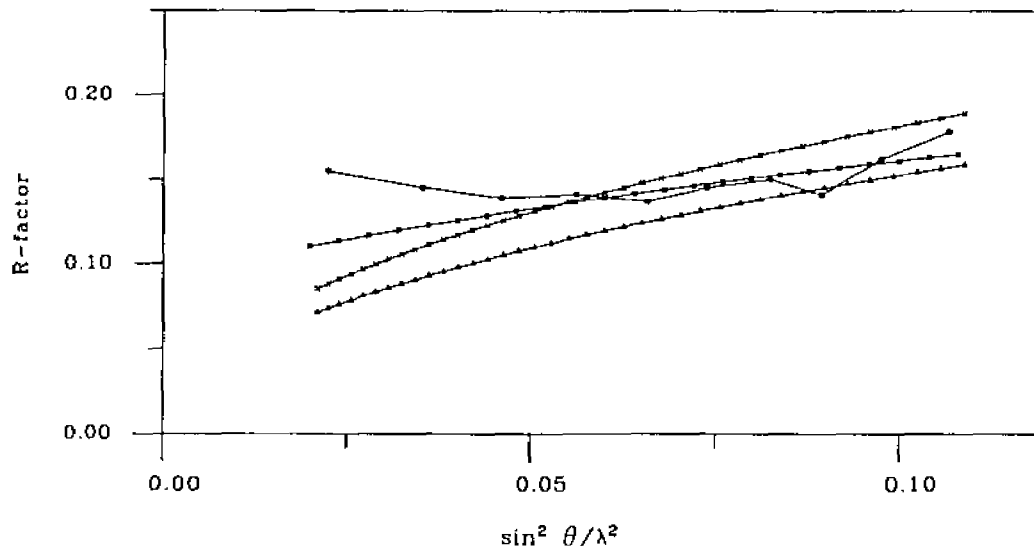


Figure 1.2.6-8. Estimation of average coordinate error from a plot of R-factor versus resolution for the reduced form of rubredoxin after refinement between 8-1.5 Å with SHELXL-93. Theoretical lines are calculated for 0.12 Å (top), 0.10 Å (bottom) and for a 97% complete model with an average error of 0.08 Å.

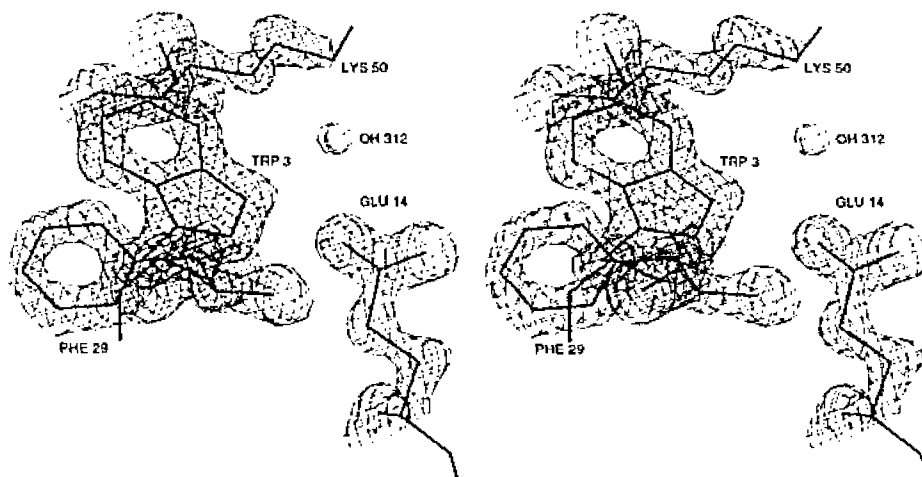


Figure 1.2.6-9. Stereo view of the electron density at 2σ from the 2Fo-Fc map in the area around the residues believed to play an important role in the thermal stability of RdPf, Trp-3 and Glu-14.

References

- Allen, F. H., Kennard, O. and Taylor, R. (1983). "Systematic Analysis of Structural Data as a Research Technique in Organic Chemistry." *Acc. Chem. Res.* **16**, 146-153.
- Engh, R. A. and Huber, R. (1991). "Accurate Bond and Angle Parameters For X-Ray Protein-Structure Refinement." *Acta. Cryst. A* **47**, 392-400.

Hendrickson, W. A. and Konnert, J. H. (1980). Bangalore, I.U.Cr. and Indian Academy of Sciences.

Luzzati, V. (1952). "Traitement Statistique des Erreurs dans la Determination des Structures Cristallines." *Acta Crystallographica* **5**, 802-810.

Sheldrick, G. M. (1993). "SHELXL-93 A Program for Structure Refinement." *University of Göttingen*

1.3 Structure Discussion

Overall folding

The overall folding (see Figure 1.3-1) of RdPf is very similar to the other rubredoxins which have been refined to high resolution (Watenpaugh, Sieker et al., 1979; Sieker, Stenkamp et al., 1986; Frey, Sieker et al., 1987; Adman, Sieker et al., 1991). An alpha carbon trace of the backbone is shown in Figure 1.3-2. The structure is made up of a 3-stranded anti-parallel β -sheet comprised of residues Ala-1 to Lys-6, Gly-9 to Glu-14 and Glu-47 to Leu-51, involving 9 hydrogen bonds (see Figure 1.3-3). The hydrogen bonding pattern between the first and second β strands is interrupted by a G1 type β bulge (Richardson, Getzoff et al., 1978) that occurs between residues Gly-9 and Tyr-10. The overall hydrogen bonding pattern is in close agreement with the pattern reported in the ^1H -

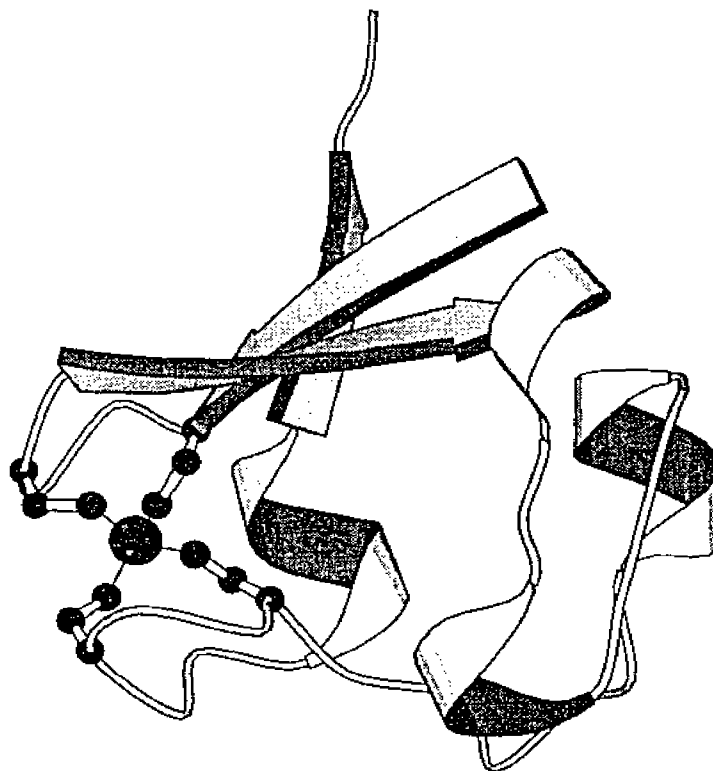


Figure 1.3-1 Molscript representation showing the overall topology of the rubredoxin from *Pyrococcus furiosus*. The iron-sulfur center is shown on the left and the 3-stranded β -sheet is shown at the top. The four helical corners are also shown.

NMR secondary structure analysis (Blake, Park et al., 1991). The hydrophobic core contains six aromatic residues, Trp-3, Tyr-10, Tyr-12, Phe-29, Trp-36 and Phe 48, as well as the hydrophobic aliphatic residue Leu-32 (see Figure 1.3-4). The aromatic residues in the hydrophobic core are nearly invariant among the known rubredoxin sequences with the exception of position 3 (this is equivalent to position 4 in all other rubredoxins due to the lack of an N-terminal methionine residue in RdPf) which is Trp in RdPf and in the rubredoxin from *Clostridium thermosaccharolyticum*, and either Tyr or Phe in all other rubredoxins. The hydrophobic aliphatic side chain at position 32 is either Leu, Ile or Val in all known rubredoxins.

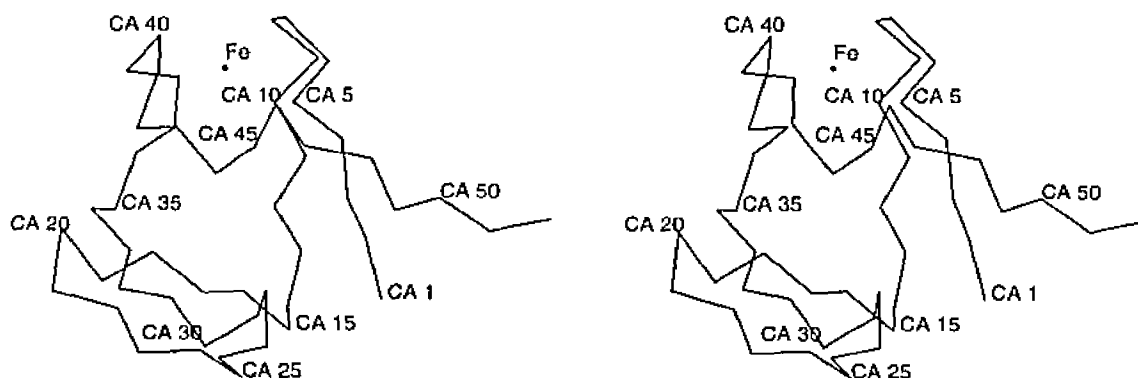


Figure 1.3-2. Stereo view of the C α trace showing the overall folding of RdPf. The iron atom is at the top of the molecule with the sheet to the right. The pendant tail on the right is the C-terminus.

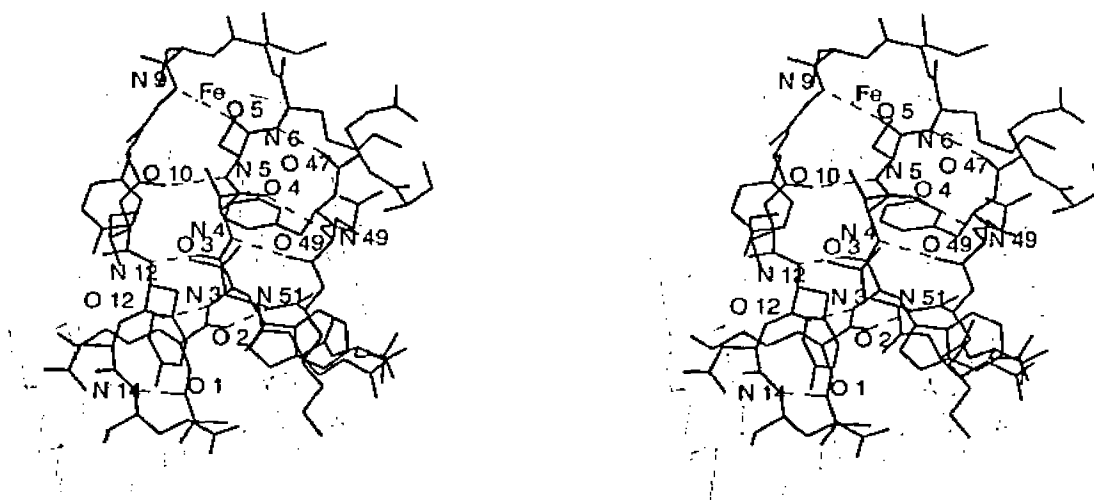


Figure 1.3-3. Stereo view of the hydrogen-bonding network in the β -sheet including the hydrogen bond, O5 \cdots N9, that defines a G-1 β -bulge.

Numerous turns are present in the RdPf structure. Of those turns stabilized by hydrogen bonds between the CO of residue n and the NH of residue $n+3$ (Venkatachalam, 1968), the two most prevalent types are the common turn and the glycine turn (Richardson, 1981; Richardson and Richardson, 1989). Overlapping common turns occur between residues 13-17, 18-22, 28-32 and 44-48. These successive turns approximate stretches of a 3_{10} helix, and were described by Watenpaugh et al. (Watenpaugh, Sieker et al., 1979) as forming helical corners in the RdCp structure. Other turns include residues 24-27 (forming a glycine turn) and 33-36 (forming a common turn). Additionally, residues Asp-13 to Asp-15 form an Asx turn (Richardson, 1981; Rees, Lewis et al., 1983), which resembles a reverse turn in which an Asp-13 side chain carboxyl oxygen forms a hydrogen bond with the amide nitrogen of Asp-15.

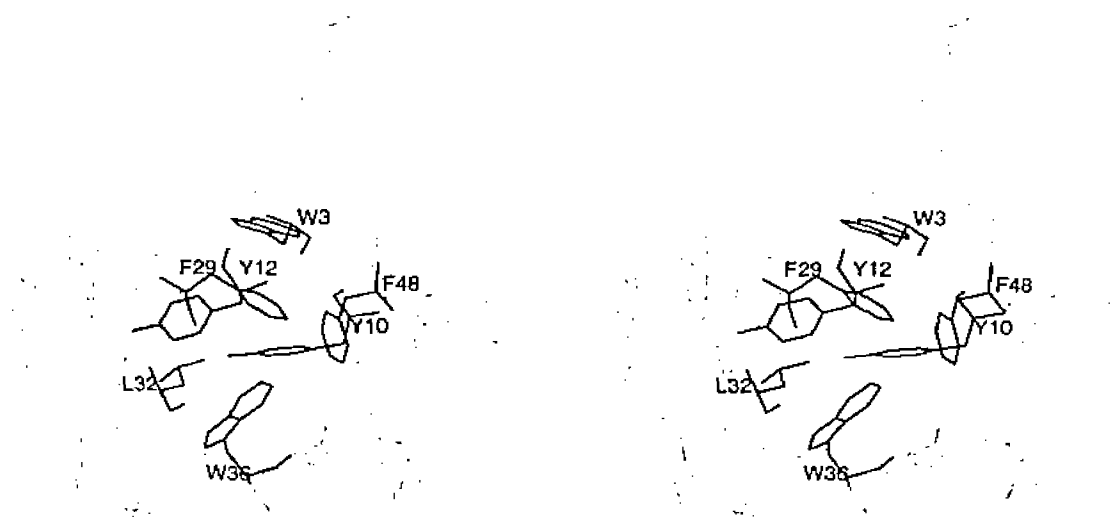


Figure 1.3-4. Stereo view of the aromatic residues and Leu 32 that form the hydrophobic core of RdPf.

Iron-sulfur environment

The sulfur atoms of the four cysteine residues ligand the iron atom with nearly equal bond lengths (see Table 1.3-2), forming an approximately tetrahedral coordination sphere. As noted by Watenpaugh et al. (Watenpaugh, Sieker et al., 1979), an approximate

two-fold axis relates two loops of the polypeptide chain, one consisting of residues 4-11, and the other of residues 37-44 (Figure 1.3-5). Furthermore, this approximate two-fold axis is reflected in the pattern of the S-Fe-S bond angles (see Table 1.3-2). Additionally, the cluster environment includes six N-H...S γ hydrogen bonds (see Table 1.3-3), two to C5S γ from I7NH and C8NH, two to C38S γ from I40NH and C41NH, one each to C8S γ from Y10NH, and, to C41S γ from A43NH. This network of hydrogen bonds also reflects the approximate two-fold symmetry about the cluster.

It is interesting to notice in Table 1.3-2 that two of the Fe-S bonds are shorter than the other two Fe-S bonds. The two longer bonds are from iron to Cys-5 S γ and to Cys-38 S γ , both of which form two hydrogen bonds with the main chain (see Table 1.3-3) and are situated toward the hydrophobic core of the protein. Additionally, these "short" and "long" bonds are consistent with the two-fold symmetry of the site (see Figure 1.3-5). Although this grouping of the iron-sulfur bond lengths is not observed in the 1.5Å refinement of the reduced form, it has been observed in other rubredoxin crystal structures which have been deposited in the Brookhaven Data Base (see Table 1.3-1) (Sieker, Stenkamp et al., 1994). There are certainly significant variations between the structures but the average values may be informative. Consideration must be given to the resolution of the refinements (8RXN 1.0Å, 7RXN 1.5Å, 6RXN 1.5Å, 5RXN 1.2Å, 4RXN 1.2Å, 1RDG 1.4Å, 1CAA 1.8Å, 1CAD 1.8Å) and to the decisions made as to the target values for the iron-sulfur distance during refinement when evaluating these observations. The average value for the Fe-S γ 5 and Fe-S γ 38 distance is 2.31 Å compared to an average of 2.26 Å for the Fe-S γ 8 and Fe-S γ 41 pair. Two of the structures in Table 1.3-1 (4RXN and 1CAA) were refined without restraints on the iron-sulfur distance, and the trend is not as obvious in 4RXN as in 1CAA. If the high resolution (i.e. $d \leq 1.2$ Å) structures only (8RXN, 5RXN and 4RXN) are considered, then the respective average values are 2.31 Å and 2.27 Å.

This trend is also observed in the coordinates refined with SHELXL-93, but is affected by the weighting scheme employed during refinement (see Table 1.3-1 and Figure 1.3-5). With modified $\frac{1}{\sigma^2}$ weights the bonds are more nearly equal than when $\frac{1}{\sigma}$ weights are used. In either case the values are not significantly different at the 2σ level, but the trend is interesting. The NH...S distances show no obvious trend, either in the refinement here using SHELXL-93 or in the deposited coordinates. This apparent trend in the distances may be resolved if more high resolution structures become available.

Table 1.3-1 Bond distances (\AA) for the iron-sulfur center for all rubredoxin coordinate sets in the Brookhaven data bank (top) and for the current refinement of RdPf with SHELXL-93 (bottom). The numbers in parentheses are the estimated standard deviations (esd) in the bond distance. For the SHELXL-93 refinement the esd's are calculated from the full correlation matrix.

	8RXN	7RXN	1RDG	6RXN	4RXN	5RXN	1CAA	1CAD	Ave. ox.
Fe-S γ 5	2.29	2.33	2.32	2.28	2.34	2.32	2.32	2.34	2.31(2)
Fe-S γ 8	2.26	2.29	2.29	2.26	2.29	2.29	2.25	2.29	2.27(2)
Fe-S γ 38	2.29	2.29	2.28	2.31	2.30	2.30	2.33	2.35	2.30(2)
Fe-S γ 41	2.26	2.27	2.27	2.23	2.25	2.25	2.25	2.29	2.25(1)
SHELXL-93 Refinement									
	Oxidized		Reduced						
	Modified $1/\sigma^2$	$1/\sigma^2$	Modified $1/\sigma^2$	$1/\sigma^2$					
Fe-S γ 5	2.289(5)	2.300(7)	2.313(14)	2.35(2)					
Fe-S γ 8	2.275(5)	2.266(8)	2.290(20)	2.31(2)					
Fe-S γ 38	2.302(5)	2.305(7)	2.352(14)	2.35(2)					
Fe-S γ 41	2.279(5)	2.270(6)	2.330(20)	2.32(2)					

Table 1.3-2. Bond distances (Å) and bond angles (degrees) in the iron-sulfur cluster after refinement of the coordinates with XPLOR or TNT.

Bond distances to iron ^a							
			Oxidized		Reduced		
			1.8Å res	1.1Å res	1.8Å res	1.5Å res	
C5S _γ	Fe		2.31	2.30	2.34	2.32	
C8S _γ	Fe		2.25	2.26	2.29	2.27	
C38S _γ	Fe		2.33	2.35	2.36	2.33	
C41S _γ	Fe		2.25	2.24	2.29	2.32	
Bond angles around Fe							
			Oxidized		Reduced		
			1.8Å res	1.1Å res	1.8Å res	1.5Å res	
C5S _γ	Fe	C8S _γ	113.4	112.6	112.2	112.6	
C5S _γ	Fe	C38S _γ	112.0	110.0	113.8	114.2	
C5S _γ	Fe	C41S _γ	102.6	102.7	104.7	103.2	
C8S _γ	Fe	C38S _γ	102.2	102.5	102.7	102.7	
C8S _γ	Fe	C41S _γ	115.1	116.6	111.4	112.7	
C38S _γ	Fe	C41S _γ	112.1	112.7	112.6	111.8	
Torsion angles about SG							
				Oxidized		Reduced	
				1.8Å res	1.1Å res	1.8Å res	1.5Å res
Fe	C5S _γ	C5Cβ	C5Cα	169°	171°	179°	175°
Fe	C8S _γ	C8Cβ	C8Cα	90°	95°	87°	96°
Fe	C38S _γ	C38Cβ	C38Cα	169°	174°	176°	175°
Fe	C41S _γ	C41Cβ	C41Cα	88°	92°	87°	96°

^aThe iron-sulfur bond distance force constant in X-PLOR was set to zero during refinement.

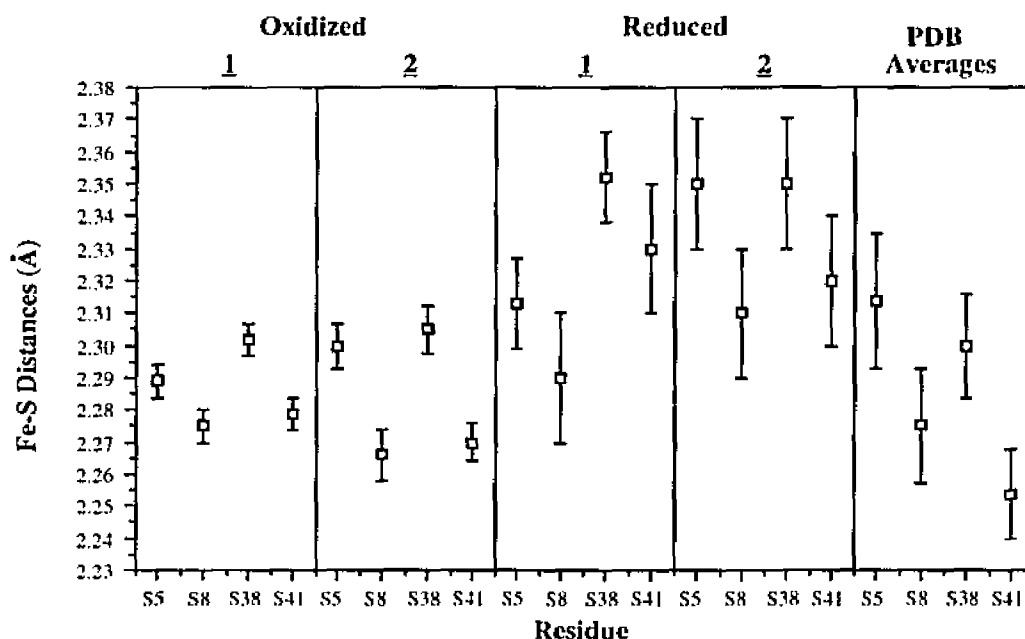


Figure 1.3-5. Plot of the bond distances between iron and sulfur showing the respective esd's. Panels labeled 1 reflect the distances and errors using the modified weighting scheme. Panels labeled 2 reflect the distances and errors using $1/\sigma^2$ weights. The esd's shown in the panel for the PDB averages are simply the esd for the average.

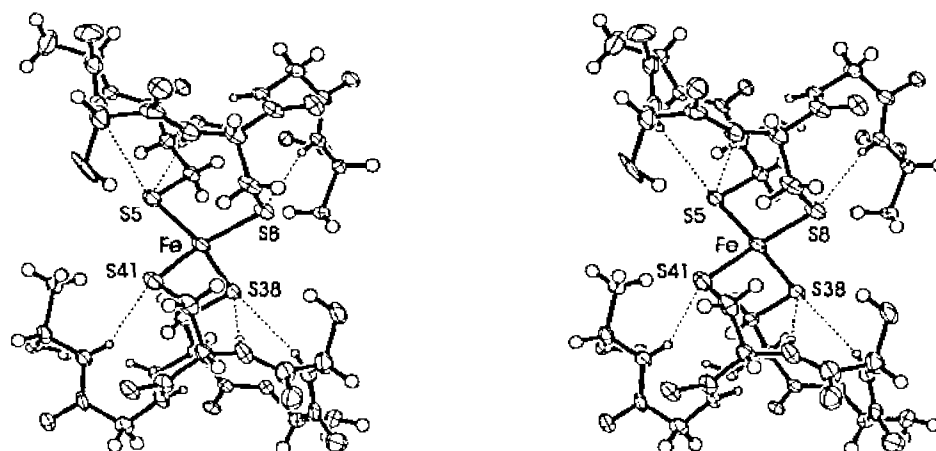


Figure 1.3-6. Stereo view of the pseudo-twofold axis around the iron-sulfur cluster. NH...S bonds are shown. Atoms are represented by 10% probability ellipsoids for clarity.

Comparison of oxidized and reduced forms

As noted above, there are few differences between the structures of the oxidized and reduced forms of RdPf, which is reflected in an rms difference in C α coordinates of 0.27Å. By minimizing the conformational changes associated with oxidation/reduction,

electron transfer reactions involving rubredoxin should be kinetically facilitated. Reduction of RdPf is accompanied by an increase in the iron-sulfur bond lengths by an average of 0.04 Å in the low resolution structures and 0.02 Å in the high resolution structures (see Table 1.3-1). The same lengthening of the iron-sulfur bonds is observed in the coordinates refined with SHELXL-93 (see Table 1.3-3). For these models the increase is 0.04 Å and 0.05 Å for the modified $\frac{1}{\sigma^2}$ and $\frac{1}{\sigma^2}$ weighted refinements respectively. It should be noted that these bond lengths were not restrained during refinement, and although the difference in bond lengths between the reduced and oxidized forms is within the estimated coordinate error, there is a consistent increase for all four of them (with the exception of Fe-S γ 38 in the 1.5 Å refinement with TNT). There is also a decrease in the average NH...S hydrogen bond distance in the 1.8 Å refinements by 0.09 Å (0.06 Å high resolution) upon reduction (see Table 1.3-4). The coordinates resulting from the SHELXL-93 refinement reflect the same shortening in the NH...S hydrogen bond distances (see Table 1.3-3). This shortening would help stabilize the negative charge introduced upon reduction, and is consistent with an observed shortening of the NH...S distances upon reduction of the oxidized form of the high-potential iron protein (Carter, Kraut et al., 1974). Although of doubtful functional relevance, the quantitatively most significant difference between oxidized and reduced RdPf in the crystals occurs at the C-terminus. The positions of the C α of Asp-53 differ by 1.38 Å between the two forms with even larger differences observed for the side chain atoms. This residue makes an intermolecular contact with a neighboring molecule; additionally there is a hydrogen bond between the amide nitrogen of Glu-52 in the reduced molecule and the carbonyl oxygen of Ile-40 in a neighboring molecule. This lattice contact is near Cys-41, which is an iron liganding residue. It is possible that reduction of the protein in the crystalline state disturbs this lattice contact, thereby accounting for the extremely fragile nature of the reduced crystal. None of the other intermolecular contacts exhibit a similar displacement.

Table. 1.3-3 *Hydrogen bond distances (Å) for sulfur to nitrogen in the coordinates refined with SHELXL-93.*

	Oxidized		Reduced	
	Modified $1/\sigma^2$	$1/\sigma^2$	Modified $1/\sigma^2$	$1/\sigma^2$
Sγ5-N7	3.564(16)	3.505(20)	3.433(27)	3.394(35)
Sγ5-N8	3.639(17)	3.616(22)	3.469(28)	3.438(33)
Sγ8-N10	3.483(12)	3.498(14)	3.417(27)	3.354(34)
Sγ38-N40	3.480(15)	3.471(19)	3.447(25)	3.448(35)
Sγ38-N41	3.605(15)	3.570(19)	3.550(27)	3.608(35)
Sγ41-N43	3.516(14)	3.505(17)	3.440(26)	3.403(31)

Comparison to other rubredoxins of known structure

The overall folding, the environment around the iron atom, the hydrophobic core and the hydrogen bonding network of RdCp, RdDs, RdDv, RdDg and RdPf are extremely similar. The rms deviation in the Cα atomic positions between RdPf (SHELXL-93) and the other rubredoxin crystal structures, RdPf (ICAA), RdCp (5RXN), RdDs (6RXN), RdDv (8RXN) and RdDg (1RDG) are 0.08Å, 0.47Å, 0.72Å, 0.56Å and 0.63Å, respectively (see Figure 1.3-6). The deviations for the Cα's of the aromatic residues in the conserved hydrophobic core are 0.06Å, 0.26Å, 0.39Å, 0.32Å and 0.26Å, respectively. Chothia and Lesk (Chothia and Lesk, 1986) have observed that the coordinate divergence, Δ, between two structures can be estimated from their fractional sequence difference, H, by the relationship:

$$\Delta = 0.40e^{1.87H} \quad (1)$$

where Δ is measured in Å. The fractional sequence differences (H) between RdPf and RdCp, RdDs, RdDv and RdDg are 0.42 (22/52), 0.55 (25/45), 0.35 (18/52), 0.35 (18/51), respectively, which correspond to calculated Δ values of 0.87Å, 1.13Å, 0.76Å and 0.77Å, respectively. Therefore, the observed structural similarities are consistent with

the general degree of sequence conservation observed in the rubredoxin family.

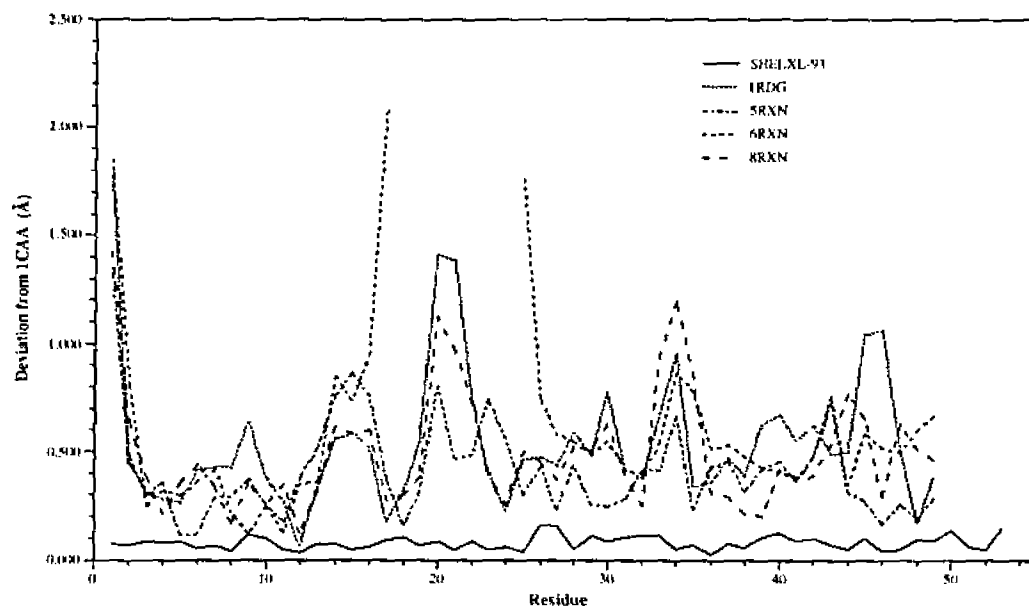


Figure 1.3-7. Plot of the root mean square differences in the backbone atoms of the various rubredoxin crystal structures after superposition on the deposited coordinates of *P. furiosus* (1CAA).

Despite the overall similarities, there are differences between the rubredoxin structures that may contribute to the thermostability of RdPf. For the purposes of comparing the RdPf structure to other rubredoxins, it is convenient to focus on three features: main chain to main chain hydrogen bonds, buried surface area, and amino acid residues unique to RdPf. In this discussion, the assumption is made that RdPf is more stable than the other rubredoxins of known structure. As mentioned in the introduction, it appears that RdPf is more stable than RdCp (Lovenberg and Sobel, 1965) and RdDg (Papavassilou and Hatchikian, 1985). No studies of the stabilities of either RdDs and RdDv have apparently been reported. Consequently, while it is reasonable to assume that RdPf is more stable than the rubredoxins from mesophilic organisms, this has not been rigorously established.

Main chain to main chain hydrogen bonds

Relative to other rubredoxins, RdPf has a more extensive network of main chain to

main chain hydrogen bonds. In RdPf, the β -sheet hydrogen bonding network extends up to residue 1, while in RdCp, RdDs, RdDv and RdDg, the hydrogen bonding network stops at residue 3. Extension of the hydrogen bonding network to the amino terminus permits formation of two additional β -sheet hydrogen bonds; one between the carbonyl oxygen of Ala-1 and the amide nitrogen of Glu-14, and the second between the carbonyl oxygen of Lys-2 and the amide nitrogen of Leu-51. Relative to mesophilic rubredoxins, the more extensively hydrogen bonded β -sheet in RdPf could enhance the stability of this structure against thermal disruption.

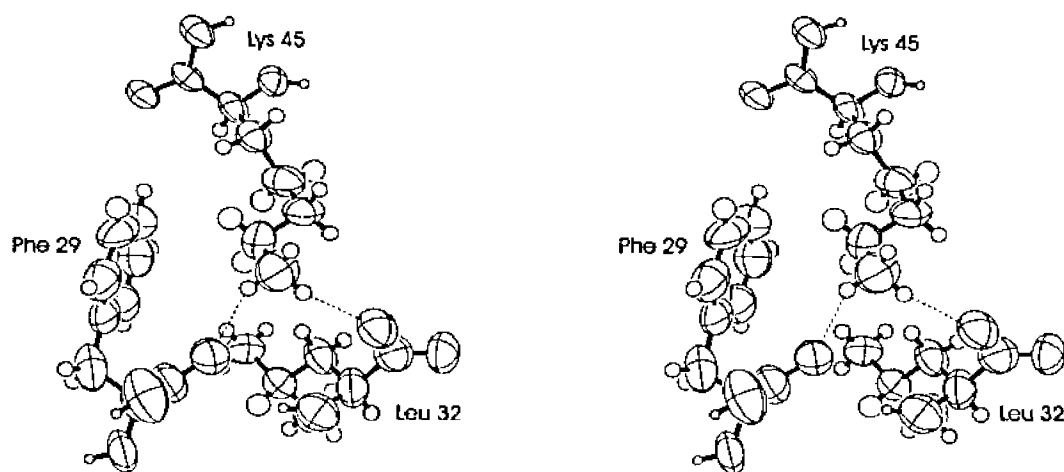


Figure 1.3-8. Side chain to main chain hydrogen bond from Lys-45, Phe-29 and Leu-32 that occurs in all known rubredoxins. Ellipsoids shown at 60% probability.

Table 1.3-3. Hydrogen bonds (Å) observed between protein atoms in RdPf

Table 13-5: Hydrogen bonds (H) observed between protein atoms in Rm

Donor	Acceptor	Oxidized (Å)		Reduced (Å)	
		1.8Å res	1.1Å res	1.8Å res	1.5Å res
Main chain to main chain					
W3N	Y12O	2.9	2.9	2.9	2.9
V4N	E49O	2.8	2.8	2.8	2.8
C5N	Y10O	2.9	2.9	2.8	2.8
K6N	E47O	2.8	2.8	2.8	2.9
G9N	C5O	2.9	2.9	3.0	3.0
Y12N	W3O	2.7	2.8	2.8	2.8
E14N	A1O	3.0	2.9	3.0	3.0
A16N	D13O	2.8	2.9	2.9	2.9
G17N	D13O	3.0	2.9	2.9	2.9
D18N	I23O	3.0	2.8	2.9	3.0
N21N	D18O	2.9	3.0	2.9	2.9
G22N	P19O	3.2	3.1	3.0	3.0
I23N	D18O	3.0	3.0	3.0	3.0
G26N	E14O	2.8	2.7	2.8	2.7
T27N	S24O	3.3	3.3	3.3	3.3
E31N	K28O	2.9	2.9	2.8	2.8
L32N	F29O	3.1	3.1	3.0	3.0
W36N	P33O	2.9	2.9	2.9	2.9
C38N	A43O	2.9	2.9	2.9	2.9
G42N	C38O	2.9	2.9	2.9	3.0
E47N	P44O	3.0	3.0	2.9	2.9
F48N	K45O	3.0	3.1	3.0	3.0
E49N	V4O	2.8	2.9	2.9	2.9
L51N	K2O	2.8	2.8	2.8	2.8
E52N	K50O	3.1	3.0	3.2	3.2
Nitrogen-Sulfur					
I7N	C5Sγ	3.52	3.47	3.34	3.39
C8N	C5Sγ	3.58	3.53	3.42	3.50
Y10N	C8Sγ	3.47	3.48	3.41	3.41
I40N	C38Sγ	3.42	3.46	3.36	3.40
C41N	C38Sγ	3.52	3.62	3.46	3.55
A43N	C41Sγ	3.49	3.48	3.50	3.43
Side chain to main chain					
Y12OH	T27O	2.6	2.6	2.6	2.7
D15N	D13Oδ1	2.9	3.0	2.9	2.9
A1N	E14Oε2	2.9	2.8	2.9	2.8
F29N	E14Oε1	2.8	2.9	2.9	2.9
T27Oγ1	S24O	2.8	2.7	2.7	2.7
E30N	E30Oε2	2.7	2.8	2.8	2.8
K45Nζ	F29O	3.0	3.0	2.8	3.2
K45Nξ	L32O	2.8	2.9	3.5	3.2
Side chain to side chain					
W3Nε1	E14Oε1	3.4	3.5	3.4	3.2
K6Nζ	E49Oε2	2.6	3.1	4.1	3.2
W36Nε1	D18Oδ2	2.9	2.9	3.0	2.9
N21Nδ2	D18Oδ1	3.0	3.2	3.2	3.0
S46Oγ	E47Oε2	2.7	2.7	2.7	2.7

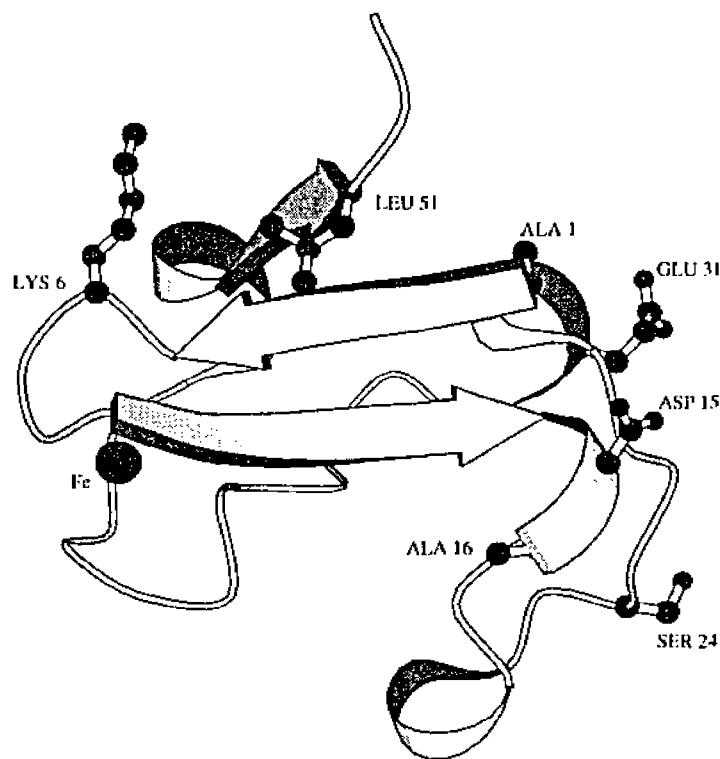


Figure 1.3-9 , Molscript drawing of the residues unique to the rubredoxin from *Pyrococcus furiosus*.

Amino acid residues unique to RdPf

The amino acid residues that are unique to RdPf are shown above in Figure 1.3-8. The interactions of the Glu-14 side chain with groups on three other residues: the amino terminal nitrogen of Ala-1; the indole nitrogen of Trp-3; and the NH group of Phe-29 (Figure 1.3-9). This could only occur in RdPf, as this is the only rubredoxin of known sequence that simultaneously contains Ala-1, Trp-3 and Glu-14. Although the Glu-14 O ϵ 1 to Trp-3 N ϵ 1 distance (3.45Å) is long for a hydrogen bond, the relative orientations of these two groups indicate that a favorable electrostatic interaction does occur. Interestingly, the only other rubredoxin known to contain a tryptophan residue at position 3 has been isolated from the thermophilic eubacterium *C. thermosaccharolyticum* (Devanathan, Akagi et al., 1969; Tanaka, Haniu et al., 1971) (Figure 1.1-3). This rubredoxin does not contain Glu-14, however. The only other rubredoxin to contain a glutamate at position 14 (isolated

from *M. elsdenii* (Figure 1.1-3)) contains neither Ala-1 nor Trp-3. The presence of this residue in RdPf that interconnects Ala-1, Trp-3, Glu-14 and Phe-29 is consistent with the idea that an increased number of salt bridges and other electrostatic interactions can enhance thermostability (Perutz, 1978). The presence of another residue uniquely found in RdPf (Lys-6) is also consistent with this idea, as this side chain is observed to form a salt bridge with the Glu-49 side chain.



Figure 1.3-10. Stereo view of the hydrogen bonds and salt bridge formed by Glu 14.

Solvent Shell

The intracellular matrix where soluble proteins function is primarily an aqueous environment and the water molecules play an important structural and functional role in defining the properties of the biological macromolecules in that environment (Timasheff, 1993). For this reason it is important to understand how a particular protein interacts with water, and since protein crystals can be up to 70% water by volume (Matthews, 1968), they provide an excellent opportunity to study that interaction. In crystals where the packing of protein molecules is close, the majority of the water has been located by x-ray crystallography (Watenpaugh, Sieker et al., 1979; Teeter and Hope, 1986; Dauter, Sieker et al., 1992). In general it can be stated that water in a protein crystal does not assume random positions (Thanki, Umrana et al., 1991).

Care must be taken when evaluating the significance of the solvent molecules found in

an x-ray crystal structure for the following reasons. The solvent in a protein crystal lattice is highly mobile and is in constant exchange with the mother liquor bathing the crystal. Macromolecular crystallographers have exploited this property in the preparation of heavy atom derivatives in order to solve the phase problem. Furthermore, the water molecules that do make contact with the protein surface have been shown by NMR to have very short residence times (between 200 to 90 ps) on the surface (Brunner, Liepinsh et al., 1993). Water may mediate protein lattice contacts and therefore have doubtful structural or biological significance. Many surface side chains can adopt multiple conformations and are themselves highly dynamic thereby providing the possibility of confusing solvent for disordered side chains. Additionally, there may be experimental errors in the data collection or model building and refinement process that produce spurious peaks in the electron density maps that may be mistaken for solvent. Finally, the solution used to crystallize the protein may be quite different than the intracellular environment with respect to pH and salt content.

Fortunately, many of these problems can be addressed during a crystallographic investigation of the solvent in a protein crystal. A very useful technique for dealing with the solvent dynamics (as well as other more nagging problems such as radiation damage) has been to collect data at liquid nitrogen temperatures using the methods of Hope (Hope, 1988; Hope, 1990). At these temperatures (approximately 100K) the solvent shell and the protein side chains are essentially static during the period of data collection. This technique has been used in the determination of the water structure in crystals of crambin (Teeter and Hope, 1986) and has enhanced the determination of the solvent structure in many other proteins. One way to address the problem of experimental error is to compare, where possible, the crystal structures of homologous proteins or of the same protein in different crystal forms (Blake, 1983; Daopin, Davies et al., 1994; Ohlendorf, 1994). In this case, the proteins would most probably be in significantly different environments and would

therefore give a relatively accurate picture of the water molecules essential to the structure.

Assuming an average protein density of 1.35 g/cm, the fractional solvent content by volume (V_p) can be calculated from the Matthews coefficient (V_m) by the following equation (Matthews, 1968):

$$V_p = 1 - \frac{1.23}{V_m}$$

For RdPf $V_m = 2.30 \frac{\text{\AA}^3}{\text{dalton}}$ corresponding to the crystals of RdPf being approximately 46% solvent by volume. This value for solvent content is very close to the average for soluble proteins of 43% (Matthews, 1968) and is consistent with the solvent content of RdCp (Watenpaugh, Seiker et al., 1973), RdDs (Sieker, Stenkamp et al., 1986), and Cts (Meyer, Gagnon et al., 1990) (only preliminary crystallographic data Cts is available to date). However, the solvent content for RdDg (Frey, Sieker et al., 1987) and RdDv (Dauter, Sieker et al., 1992) is unusually low at approximately 27%.

The crystal structures of the oxidized and reduced forms of RdPf contain 132 and 102 solvent oxygen atoms respectively. This accounts for nearly all of the solvent in the asymmetric unit of the crystals. In the oxidized form, 77 of the 132 solvents form 102 hydrogen bonds with the protein (see Table 1.3-5). The remaining 55 solvent molecules make contact with other solvent oxygens only. Of the 132 solvent oxygens, 7 have a temperature factor greater than 50\AA^2 with 3 of these in contact with the protein. One hundred twenty solvent oxygens were located in the crystal structure of the reduced rubredoxin. A total of 64 of these are within hydrogen bonding distance of the protein surface and form 88 hydrogen bonds (see Table 1.3-6). The remainder of the solvent forms contacts to other solvent molecules only. Only 9 of the solvents have $B's > 50\text{\AA}^2$, 2 of which are bonded to the protein.

The majority of the lattice contacts between protein molecules are mediated by solvent molecules. There are only two direct lattice contacts between adjacent molecules

(O Ser-46 to C α Gly-22 and N Asp-53 to O Ile-40). The other 12 lattice contacts, N ζ Lys-2 to O ϵ 1 Glu-47, N ζ Lys-2 to O Lys-6, O Gly-9 to O γ Ser-46, O Asp-20 to O ϵ 1 Glu-30, O Gly-22 and O Ser-24 to O Glu-47, O Ile-40 to O Glu-52, O Cys-41 and O Lys-50 to O ϵ 1 Glu-49, O Gly-42 to O ϵ 1 Glu-49 and O Phe-48 and O Ser-46 to O Gly-22, are all mediated by solvent molecules.

Perhaps the most interesting question that can be asked about the solvent shell of any protein in a crystalline lattice involves the structural significance of the water bound to the protein. In an effort to address this question, the solvent shell of the oxidized form was compared to the reduced form. It was found that 41 of the solvents are common to both forms. Of these 41, there are 28 that make direct contact to the protein surface and another 11 that contact these surface bound solvents but do not contact the protein. The comparison was then extended to the rubredoxin crystal structures deposited in the Protein Data Bank. The criteria for sameness in this comparison was that the solvents lie within 0.9Å of a solvent on the oxidized form and that it be within hydrogen bonding distance of the analogous residue in at least one other crystal structure. The results are summarized in Table 1.3-4 below. It was found that 20 solvent molecules satisfied this criteria. The majority of these occurred in at least two or more other structures. The least homology was observed with RdDs, which is not surprising considering that RdDs is unique from all the other rubredoxins whose crystal structure is known by virtue of missing a 7 residue stretch in the middle of the sequence (see Figure 1.3-7). It is interesting to notice that virtually all of these conserved water molecules are associated with distinct areas of secondary structure. Conserved waters lie along the edge of the β -sheet consistent with the observation by Thornton et al. (Thanki, Umrania et al., 1991) that the edges of a β -sheet are a common site for water. Other common waters are associated with the helical turns and are possibly stabilizing elements in these structures. Another set of conserved waters group around residues 28 to 32. This section of the structure can best be described as

random coil, but it passes between two helical turns, and water here serves to increase the contact area and potentially lend increased structural stability.

These results would seem to support the hypothesis that solvent does not occupy merely random positions in a protein crystal. There are in fact a number of other similar observations. Blake et al. observed a high degree of commonalty between different crystal forms of lysosymes and between homologous lysosymes (Blake, 1983). Three independent investigations of the crystal structure of interleukin 1 β contain between 30 and 74 common solvent molecules (Ohlendorf, 1994). A comparison of two crystal structures of transforming growth factor- β 2 indicate similar tendencies for water to be in, what the authors call, non random positions within the unit cell (Frey, 1994).

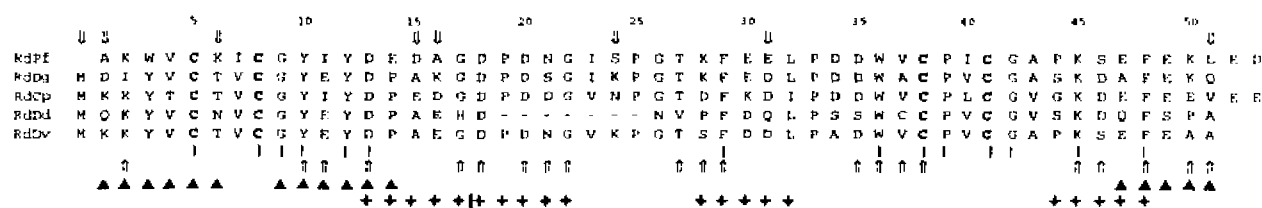


Figure 1.3-11. Sequences of RdPf and the four other rubredoxins whose x-ray crystal structure is known. The cystien residues are shown in bold face, ↑ shows the location of the conserved solvent molecules, ↓ denotes residues unique to RdPf and I marks residues that are strictly conserved in all known rubredoxin sequences. The symbols along the bottom denote elements of secondary structure; ▲ represents strands of the β -sheet and + represents the helical corners.

Table 1.3-4. Solvent atom protein contacts in RdPf that occur in at least one other rubredoxin crystal structure deposited in the PDB. Only solvent oxygens that are within 0.9 Å of a RdPf solvent oxygen and make a contact to the same site are listed. Solvent oxygens with numbers between 101-199 are common to both the oxidized and reduced form of RdPf.

Solvent Number	Residue	Atom	1RDG RdDg	5RXN RdCp	6RXN RdDs	8RXN RdDv
101	TYR 10	OH	✓	✓		✓
	GLY 17	O	✓	✓		✓
	ASP 13	N	✓	✓		✓
104	CYS 38	O	✓	✓		
105	LYS 45	N	✓	✓		
	TRP 36	O	✓			
106	VAL 37	O	✓			
	ASP 18	OD2	✓			
107	LYS 2	N		✓	✓	✓
	LEU 51	O		✓	✓	✓
108	VAL37	N	✓	✓		✓
109	ILE 11	O		✓		
110	ASP 20	N		✓		
111	ASP 13	OD1			✓	
112	TYR 10	OH		✓		✓
115	LYS 50	N	✓			✓
117	GLY 22	N	✓			
122	THR 27	OG1	✓	✓		
301	LYS 45	O				✓
	PHE 28	O	✓			
308	GLY 17	O	✓	✓		✓
314	SER 46	O	✓		✓	
	PHE 48	O	✓		✓	
321	LYS 28	N	✓	✓		✓
327	ASP 13	OD2				✓
348	ASN 21	OD1		✓		
353	ASP 35	O		✓		✓

Table 1.3-5. Solvent oxygens within H-bonding distance of the oxidized form of RdPf. # indicates a residue on a symmetry related molecule and a parenthesis encloses the number solvent for solvents with $B > 50 \text{ \AA}^2$.

Solvent Number	Atom Type	Residue Type	Residue Number	Bond Distance	Solvent Number	Atom Type	Residue Type	Residue Number	Bond Distance
101	O	GLY	17	2.87	307	O	ASP	15	2.83
	OH	TYR	10	2.85	308	O	GLY	17	2.79
	N	ASP	13	2.94	313	O	CYS	8	2.68
102	OG	SER	24	2.68	314	O	SER	46	2.87
	O	PRO	25	2.70		O	PHE	48	2.87
103	O	GLY	9	2.75	315	OE2	GLU	30	2.87
104	O	CYS	38	2.74		OE1	GLU	14	2.61
105	O	ASP	34	2.79	316	O	ALA	16	2.86
	O	TRP	36	2.65	317	OD2	ASP	15	3.08
	N	LYS	45	3.05		OD2	ASP	13	2.86
106	OD2	ASP	18	2.78	318	O	ASN	21	2.84
	O	VAL	37	2.73	319	N	ASP	34	3.03
107	N	LYS	2	2.95	321	N	LYS	28	2.90
	O	LEU	51	2.77	325	OD1	ASP	15	2.71
108	N	VAL	37	3.13	326	O	LYS	6	2.76
109	O	ILE	11	2.62	327	OD2	ASP	13	2.72
110	OD1	ASP	18	2.80	328	OD1	ASP	15	2.60
	N	ASP	20	2.87	329	O	LEU	32	2.91
111	N	ALA	1	2.81	331	OD2	ASP	35	2.74
	O	ALA	1	3.22	334	O	GLY	22	3.20
	OD1	ASP	13	2.89	335	NZ	LYS	2	2.84
	OD2	ASP	15	3.11	336	NZ	LYS	2	3.01
112	OD1	ASP	18	2.93	337	OE1	GLU	52	2.64
	OH	TYR	10	2.64		O	LYS	50	2.97
113	O	GLY	42	2.55	339	NZ	LYS	28	3.02
114	OD1	ASN	21	3.29		OE2	GLU	14	2.76
115	N	LYS	50	3.06	341	O	ASP	20	2.92
116	O	ASP	20	3.29		OD1	ASN	21	2.95
117	O	GLY	22	2.82	342	O	PRO	39	2.85
	N	SER	24	3.08	346	OE2	GLU	30	2.76
118	OG	SER	24	2.80	347	OE2	GLU	14	2.95
119	OG1	THR	27	2.92	351	CE	LYS	2	3.22
	N	SER	24	3.26		NZ	LYS	2	3.08
120	OE1	GLU	30	2.74	353	O	ASP	35	2.78
	NZ	LYS	50	2.54	354	N	ALA	1	3.09
121	O	CYS	41	2.82		OD1	ASP	15	2.95
122	OG1	THR	27	2.74	356	O	GLU	31	2.68
123	N	ALA	1	2.91	362	OD2	ASP	20	2.88
126	O	ASP	35	2.88	363	NZ	LYS	2	3.09
127	O	ILE	40	3.00	364	OXT	ASP	53	2.80
128	O	GLY	42	2.86	367	CE	LYS	6	3.26
129	ND2	ASN	21	2.97	369	OE1	GLU	52	2.95
130	O	ASP	20	2.93	370	OD1	ASN	21	3.14
301	O	LYS	45	2.78	371	CE	LYS	6	3.24
	O	PHE	48	2.94	372	O	PRO	39	3.06
302	O	ASP	#34	2.80	376	NZ	LYS	6	2.82
303	O	PRO	19	2.96	374	NZ	LYS	50	2.55
304	OD1	ASP	34	2.68	(379)	O	LEU	51	3.24
	NZ	LYS	45	2.80	383	OD2	ASP	15	2.91
305	OD1	ASP	20	2.63	(389)	O	ILE	7	2.91
306	NZ	LYS	#45	3.20	(399)	OE1	GLU	30	2.93

Table 1.3-6. Solvent oxygens within H-bonding distance of the reduced form of RdPf.

Solvent Number	Atom Type	Residue Type	Residue Number	Bond Distance	Solvent Number	Atom Type	Residue Type	Residue Number	Bond Distance
101	OH	TYR	10	2.86	309	O	PRO	19	2.84
	O	GLY	17	2.81	311	OD2	ASP	15	2.85
	N	ASP	13	2.90		OD2	ASP	13	2.61
102	OG	SER	24	2.72	312	OE2	GLU	30	2.79
	O	PRO	25	2.64		OE1	GLU	14	2.63
103	O	GLY	9	2.76	315	OD2	ASP	13	2.59
104	O	CYS	38	2.77	316	O	GLY	22	3.16
105	O	ASP	34	2.93	317	O	CYS	8	2.76
	N	LYS	45	3.05	321	NZ	LYS	2	3.02
	O	TRP	36	2.72	322	OD2	ASP	15	2.62
106	O	VAL	37	2.71	323	OE2	GLU	14	2.69
	OD2	ASP	18	2.77	326	O	ASP	20	2.90
107	N	LYS	2	2.99		OD1	ASN	21	2.79
	O	LEU	51	2.73	327	OE1	GLU	52	3.24
108	N	VAL	37	3.27		OD1	ASP	53	2.88
109	O	ILE	11	2.63		OXT	ASP	53	2.71
110	OD1	ASP	18	2.82	329	N	ALA	1	3.14
	N	ASP	20	2.83		OD1	ASP	15	2.97
111	N	ALA	1	2.99	331	NZ	LYS	2	2.96
	OD1	ASP	13	2.66	332	CE	LYS	2	3.16
	OD2	ASP	15	2.76	334	O	ASN	21	2.84
112	OD1	ASP	18	3.27	336	OD1	ASN	21	3.19
	OH	TYR	10	2.59		ND2	ASN	21	2.98
113	O	GLY	42	2.58		O	PRO	39	2.96
114	OD1	ASN	21	2.99	337	OE2	GLU	52	3.12
115	N	LYS	50	2.97		O	LYS	50	2.98
116	O	ASP	20	3.11	341	OD1	ASN	21	2.77
117	O	GLY	22	2.87	344	NZ	LYS	2	2.87
	N	SER	24	3.17	345	OE2	GLU	49	2.48
118	OG	SER	24	2.72	349	OE1	GLU	47	2.72
119	OG1	THR	27	2.74	351	OD1	ASP	20	3.10
120	OE1	GLU	30	2.71	353	N	ASP	53	2.65
121	O	CYS	41	2.54	358	O	ASP	35	2.57
122	OG1	THR	27	2.53	361	CG	GLU	30	3.17
123	N	ALA	1	2.85		OE1	GLU	30	3.16
124	OD1	ASP	15	2.57	365	OE1	GLU	31	2.58
126	O	ASP	35	3.15	368	OE1	GLU	31	3.27
127	O	ILE	40	2.85	374	OD2	ASP	34	3.26
128	O	GLY	42	2.62		N	ASP	34	2.85
130	O	ASP	20	2.60	384	O	ASP	34	3.22
303	O	GLY	17	2.82	386	OD1	ASP	20	2.92
305	O	ALA	16	2.59		OD2	ASP	20	2.75
307	O	ILE	7	2.66	387	O	GLU	31	3.03
308	OD1	ASP	15	2.83	388	OD1	ASP	20	2.49

References

- Adman, E. T., Sieker, L. C. and Jensen, L. H. (1991). "Structure of Rubredoxin from *Desulfovibrio vulgaris* at 1.5 Å Resolution." *Journal of Molecular Biology* **217**, 337-352.
- Blake, C. C. F. (1983). "X-ray Studies of Water in Crystals of Lysozyme." *J. Mol. Biol.* **167**, 693-723.
- Blake, P. R., Park, J.-B., Bryant, F. O., Aono, S., Magnuson, J. K., Eccleston, E., Howard, J. B., Summers, M. F. and Adams, M. W. W. (1991). "Determinants of Protein Hyperthermostability: Purification and Amino Acid Sequence of Rubredoxin from the Hyperthermophilic Archaeobacterium *Pyrococcus furiosus* and Secondary Structure of the Zinc Adduct by NMR." *Biochemistry* **30**, 10885-10895.
- Brunne, R. M., Liepinsh, E., Otting, G., Wüthrich, K. and van Gunsteren, W. F. (1993). "A Comparison of Experimental Residence Times of Water Molecules Solvating the Bovine Pancreatic Trypsin Inhibitor with Theoretical Model Calculations." *J. Mol. Biol.* **231**, 1040-1048.
- Carter, C. W., Kraut, J., Jr., Freer, S. T. and Alden, R. A. (1974). "Comparison of Oxidation-Reduction Site Geometries in Oxidized and Reduced *Chromatium* High Potential Iron Protein and Oxidized *Peptococcus aerogenes* Ferredoxin." *Journal of Biological Chemistry* **249**, 6339-6346.
- Chothia, C. and Lesk, A. M. (1986). "The relation between the divergence of sequence and structure in proteins." *The EMBO Journal* **5**, 823-826.
- Daopin, S., Davies, D. R., Schlunegger, M. P. and Grütter, M. G. (1994). "Comparison of Two Crystal Structures of TGF-β2: the Accuracy of Refined Protein Structures." *Acta. Cryst.* **D50**, 85-92.
- Dauter, Z., Sieker, L. C. and Wilson, K. S. (1992). "Refinement of Rubredoxin from *Desulfovibrio vulgaris* at 1.0Å with and without Restraints." *Acta Cryst.* **B48**, 42-59.
- Devanathan, T., Akagi, J. M., Hersh, R. T. and Himes, R. H. (1969). "Feredoxin from two thermophilic clostridia." *Journal of Biological Chemistry* **244**, 2846-2853.
- Frey, M. (1994). "Water Structure Associated with Proteins and its Role in Crystallization." *Acta. Cryst.* **D50**, 663-666.
- Frey, M., Sieker, L., Payan, F., Haser, R., Bruschi, M., Pepe, G. and LeGall, J. (1987). "Rubredoxin from *Desulfovibrio gigas* A Molecular Model of the Oxidized Form at 1.4Å Resolution." *Journal of Molecular Biology* **197**, 525-541.
- Hope, H. (1988). "Cryocrystallography of Biological Macromolecules: a Generally Applicable Method." *Acta Crystallographica* **B44**, 22-26.
- Hope, H. (1990). "Crystallography of Biological Macromolecules At Ultra-Low Temperature." *Annual Review of Biophysics and Biophysical Chemistry* **19**, 107-126.
- Lovenberg, W. and Sobel, B. E. (1965). "Rubredoxin: A new electron transfer protein from *Clostridium pasteurianum*." *Proc. Natl. Acad. Sci. USA* **54**, 193-199.
- Matthews, B. W. (1968). "Solvent Content of Protein Crystals." *Journal of Molecular Biology* **33**, 491-497.

- Meyer, J., Gagnon, J., Sieker, L. C., van Dorsselaer, A. and Moulis, J.-M. (1990). "Rubredoxin from *Clostridium thermosaccharolyticum*: Amino Acid Sequence, Mass-Spectrometric and Preliminary Crystallographic Data." *Biochemical Journal* **271**, 839-841.
- Ohlendorf, D. H. (1994). "Accuracy of Refined Protein Structures. II. Comparison of Four Independently Refined Models of Human Interleukin 1 β ." *Acta. Cryst.* **D50**, 808-812.
- Papavassilou, P. and Hatchikian, E. C. (1985). "Isolation and characterization of a rubredoxin and a two-[4Fe-4S] ferredoxin *Thermodesulfobacterium commune*." *Biochim. Biophys. Acta* **810**, 1-11.
- Perutz, M. F. (1978). "Electrostatic effects in proteins." *Science* **201**, 1187-1191.
- Rees, D. C., Lewis, M. and Lipscomb, W. N. (1983). "Refined crystal structure of carboxypeptidase A at 1.5Å resolution." *J. Mol. Biol.* **168**, 367-387.
- Richardson, J. S. (1981). The Anatomy and Taxonomy of Protein Structure. Advances in Protein Chemistry, 34, Academic Press, New York; pp. 167-339.
- Richardson, J. S., Getzoff, E. D. and Richardson, D. C. (1978). "The β -bulge: A common small unit of nonrepetitive protein structure." *Proc. Natl. Acad. Sci. USA* **75**, 2574-2578.
- Richardson, J. S. and Richardson, D. C. (1989). Principles and patterns of protein conformation. In *Prediction of Protein Structure and the Principles of Protein Conformation* (Editor), 1-98. Plenum, New York.
- Sieker, L. C., Stenkamp, R. E., Jensen, L. H., Prickril, B. and LeGall, J. (1986). "Structure of rubredoxin from the bacterium *Desulfovibrio desulfuricans*." *FEBS* **208**, 73-76.
- Sieker, L. C., Stenkamp, R. E. and LeGall, J. (1994). "Rubredoxin in Crystalline State." *Method Enzymol.* **243**, 203-216.
- Tanaka, M., Haniu, M., Matsueda, G., Yasunobo, K. T., Himes, R. H., Akagi, J. M., Barnes, E. M. and Devanathan, T. (1971). "The Primary Structure of the *Clostridium tartarivorum* Ferredoxin, a Heat-stable Ferredoxin." *Journal of Biological Chemistry* **246**, 3953-3960.
- Teeter, M. M. and Hope, H. A. (1986). "Progress in the Water Structure of the Protein Crambin by X-Ray Diffraction at 140 K." *Ann. NY Acad. Sci.* **482**, 163-165.
- Thanki, N., Umrana, Y., Thornton, J. M. and Goodfellow, J. M. (1991). "Analysis of Protein Main-chain Solvation as a Function of Secondary Structure." *J. Mol. Biol.* **221**, 669-691.
- Timasheff, S. N. (1993). "The Control of Protein Stability and Association by Weak Interactions with Water: How do Solvents Affect These Processes?" *Ann. Rev. Biophys. Struct.* **22**, 67-97.
- Venkatachalam, C. M. (1968). "Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units." *Biopolymers* **6**, 1425-1436.
- Watenpugh, K. D., Sieker, L. C., Herriot, J. R. and Jensen, L. H. (1973). "Refinement of the model of a protein: Rubredoxin at 1.5Å resolution." *Acta Crystallogr.* **B29**, 943.
- Watenpugh, K. D., Sieker, L. C. and Jensen, L. H. (1979). "The Structure of Rubredoxin at 1.2Å Resolution." *Journal of Molecular Biology* **131**, 509-522.

1.4 Possible Determinants of Thermostability

Thermal stability in proteins arises from certain structural features present in thermophilic analogs of mesophilic proteins that are not present in the protein from the mesophilic organism. Many hypotheses have been put forward regarding general features which impart stability to one amino acid sequence, relative to another amino acid sequence which adopts the same fold. Klibanov et al. found that deamination is a major factor in the irreversible deactivation of lysozyme at 100° C and postulated that this may be true for many other proteins as well (Ahern and Klibanov, 1985). Therefore, amino acid substitutions that reduce the rate of deamination (i.e., Asn to Gln, Ile or Thr as well as Asp to Glu) should tend to increase stability. In a study of bacterial ferredoxins and of haemoglobin A2, Perutz concluded that salt bridges between residues close to the amino terminus and residues near the carboxy terminus were largely responsible for the increased stability of the thermophilic protein and he postulated that electrostatic forces dominated protein structure (Perutz and Raidt, 1975; Perutz, 1978). In a separate study, the analysis of 34 proteins and of four peptide analogs suggested that the interaction of aromatic groups may have a stabilizing influence on protein structure (Burley and Petsko, 1985). Argos et al. examined thermophilic and mesophilic molecules of lactate dehydrogenase, ferredoxin and glyceraldehyde-3-phosphate dehydrogenase and concluded that Gly→Ala, Ser→Ala, Ser→Thr, Lys→Arg, and Asp→Glu amino acid substitutions were prevalent in the thermophilic equivalents of mesophilic proteins (Argos, Rossmann et al., 1979). This suggested that stability can be increased through the incorporation of many small changes that tend to stabilize an α -helix and tend to increase the internal hydrophobicity of the protein while decreasing its external hydrophobicity. The studies cited above are not intended to be a comprehensive listing, and studies of mechanisms which may have a direct bearing on the thermal stability of rubredoxin will be discussed below.

The structure of very few proteins from thermophilic organisms have been determined by x-ray crystallography. The structure of phosphoglycerate kinase (PGK) from the moderate thermophile *Bacillus stearothermophilus* has been solved at 1.65Å and compared to its mesophilic counterpart in yeast (Davies, Gamblin et al., 1993). The midpoint for the thermal unfolding transition (T_m) as determined by differential scanning calorimetry for the mesophilic and thermophilic forms of PGK differ by less than 15° C with the mesophilic form unfolding at 53° C and the thermophile unfolding at 67° C. This corresponds to a $\Delta\Delta G=5\text{kcal/mol}$ between the two forms for the unfolding process. The authors point to 12 ion pairs in the thermophilic structure that are not present in the PGK from yeast as a likely source of the increase in stability. This is consistent with the findings of Perutz (Perutz, 1978) in the study of ferredoxin and with the observations of Walker (Walker, Wonacott et al., 1980) in dehydrogenase. Another possible factor the authors feel may contribute to the increased thermostability are charged amino acid substitutions that presumably stabilize the α -helices (Argos, Rossmann et al., 1979; Nicholson, Becktel et al., 1988).

The crystal structure of malate dehydrogenase from the thermophilic bacterium *Thermus flavus* has (PDB listing 1BMD) been solved at 1.9Å resolution (Kelly, Nishiyama et al., 1993). This enzyme is fully active at 90° C while the mesophilic counterparts are inactivated at approximately 50° C. The presence of four ion pairs per monomer that do not occur in the mesophilic analog are thought to be the major contributors to the increased thermostability. Amino acid substitutions likely to stabilize an α -helix do not appear to contribute because the majority of the amino acid substitutions found in the helices of this enzyme tend to be destabilizing. Similarly, the pattern of hydrophobic or charged amino acid substitutions for polar groups found here tend to be destabilizing.

The coordinates of three other crystal structures of proteins from thermophilic organisms appear in the Brookhaven Protein Data Bank. These include the oxidoreductase,

3-isopropylmalate dehydrogenase from *Thermus thermophilus* (1IPD), a β -amylase, endo-1,4- β -D-gluconase from *Thermomonospora fusca* (1TML) and the oxidoreductase, *holo*-D-glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus* (1GDI). The stability of this particular 3-isopropylmalate dehydrogenase (1IPD) apparently arises from increased hydrophobic contacts between the subunits of the dimer (Imada, Sato et al., 1991). The structure of the β -amylase (1TML) is not discussed in terms of its stability but only in terms of the catalytic activity (Spezio, Wilson et al., 1993). Salt bridges appear to stabilize the tertiary and quaternary structure of glyceraldehyde-3-phosphate dehydrogenase (Skarzynski, Moody et al., 1987).

Two studies involving the stability of hyperthermophiles have been reported, one involving the pressure stabilization (Hei and Clark, 1994) and a differential scanning calorimetry study of proteins from *P. furiosus* (Klump, Adams et al., 1994). Hei et al. studied the effect of high pressure (50.7 MPa, 500 atm) of the hydrogenases from the mesophile *Methanococcus maripaludis*, the moderate thermophile *Methanococcus thermolithotrophicus* and the extreme thermophiles *Methanococcus janaschii* and *Methanococcus igneus*, of the Glyceraldehyde-3-phosphate dehydrogenases from the mesophile *Saccharomyces cerevisiae* and the hyperthermophile *P. furiosus* as well as the rubredoxin from *P. furiosus* and found that the proteins from the extreme thermophiles and from the hyperthermophile were more stable at high pressure. The rubredoxin from *P. furiosus*, which was more rapidly inactivated 110° C by high pressure (50.7 MPa) than at 1.0 MPa, was the only exception. GAPDH from *Saccharomyces cerevisiae* was not effected by pressure and the hydrogenases from the mesophilic and moderately thermophilic bacterium were destabilized by pressure.

The differences in the effect of pressure on stability can be explained in terms of the volume change (ΔV_{RD}), including the solvent volume, associated with the reversible deactivation process. High pressure will tend to stabilize the state which occupies the least

volume. Kauzmann (Kauzmann, 1959) suggested that the ΔV associated with the transfer of a hydrophobic group to an aqueous environment is negative, but more recent studies done at high pressure suggest the opposite to be true (Hvidt, 1975; Heremans, 1980). Studies on small molecules have shown that the volume change associated with the interaction between charged groups and the neighboring solvent will result in a negative volume change (Hamann, 1980).

GAPDH is a tetrameric enzyme with hydrophobic contacts between the subunits and increased hydrophobicity in these contact areas may be responsible in part for the increased thermostability for the hyperthermophilic enzyme. We and others have previously postulated that increased electrostatic interactions may in part be responsible for the stability of the rubredoxin from *P. furiosus* (Blake, Park et al., 1991; Day, Hsu et al., 1992; Bradley, Stewart et al., 1993). Therefore, if the deactivation process involves the exposure of hydrophobic groups to solvent, resulting in a positive change in volume, then the native state may be preferred under high pressure, as is the case in GAPDH. If, however, the stability arises primarily from electrostatic interactions, then the structure would be expected to be destabilized by high pressure, as is observed in rubredoxin.

Klump et al. conducted a quantitative thermodynamic study on the ferredoxin from *Thermotoga* and on three proteins from *P. furiosus*: ferredoxin, glutamate dehydrogenase and rubredoxin using differential scanning calorimetry to monitor the unfolding process. The three proteins all unfolded at the same temperature of 113° C. Privalov postulated that at 112° C the hydrophobic interactions cease to have a stabilizing effect on a proteins structure (Privalov and Khechinashvili, 1974). Therefore, any stabilization arising from electrostatic interactions in these proteins is not sufficient to overcome the loss of hydrophobic stabilization at this temperature. The reduced form of the rubredoxin was denatured at 102° C which may come somewhat as a surprise in light of the similarity to the oxidized form. Substitution of Zn for Fe in rubredoxin increased the stability by

11° C, again coming somewhat as a surprise considering the similarity to the native form as determined by NMR (Blake, Day et al., 1992; Blake, Park et al., 1992). The unfolding temperature for the ferredoxins was not dependent on the oxidation state of the metal center.

Buried surface area

The contribution of hydrophobic interactions, ΔG_H , to protein stability has been characterized by analysis of the amount and type of protein surface area that becomes buried during protein folding (Chothia and Lesk, 1986). In general, ΔG_H is expressed as the product of two terms: (a) the surface area of the region buried from exposure to water during folding and (b) a surface free energy term. In a method developed by Eisenberg and McLachlan (Eisenberg and McLachlan, 1986), ΔG_H may be expressed as a sum involving the difference in solvent accessible surface area (Richards, 1977) of the i th atom, A_i , between the folded and unfolded states, and the surface free energy, $\Delta\sigma_i$, for each atom type:

$$\Delta G_H = \sum_i \Delta\sigma_i [A_i(\text{folded}) - A_i(\text{unfolded})] \quad (2)$$

where the sum is over all atoms i . The surface free energies of transfer between the protein interior and the aqueous phase for different atom types are (in cal \AA^{-2}) $\Delta\sigma(\text{C}) = 18$, $\Delta\sigma(\text{N/O}) = -9$, $\Delta\sigma(\text{O}^-) = -37$, $\Delta\sigma(\text{N}^+) = -38$, $\Delta\sigma(\text{S}) = -5$ (Eisenberg, Wesson et al., 1989). Accessible surface areas were calculated from rubredoxin coordinates with the program ACCESS (Lee and Richards, 1971), using van der Waals radii for various atom types taken from Chothia (Chothia, 1976). The accessible surface areas of atoms in a hypothetical unfolded state are taken from Eisenberg et al. (Eisenberg, Wesson et al., 1989). The results of these calculations for the five different rubredoxin structures are presented in Table 4. RdPf exhibits the greatest calculated contribution to protein stability from these hydrophobic interactions, as a consequence of burying the greatest amount of

nonpolar (carbon) surface area. The difference between ΔG_H for RdPf and some of the other rubredoxin structures, most noticeably RdCp, is not large, however, suggesting that other factors must also contribute to the thermostability of RdPf. There does not appear to be anything unusual about the surface areas exposed to water in the folded rubredoxin structures, as the total areas are close to those anticipated for water-soluble, globular proteins of comparable molecular weight, and the fraction of the exposed surface contributed by carbon atoms is also typical for water-soluble, globular proteins (Miller, Janin et al., 1987). The fraction of the buried surface area that is nonpolar in RdPf (~70%) is greater than that generally observed for water-soluble, globular proteins (~58%) (Miller, Janin et al., 1987), but this fraction does not vary significantly between the different rubredoxin structures.

Table 1.4-1. Buried and exposed accessible surface areas for rubredoxins

	A_{total} (\AA^2) ^a	A_{NP} (\AA^2) ^b	A_{buried} (\AA^2) ^c				ΔG_H^d (kcal/mole)
			C	N/O	O ⁻ /N ⁺	S	
RdPf	3398	1884	3637	1223	197	106	-46.5
RdCp	3572	1917	3580	1178	195	129	-45.9
RdDg	3325	1809	3438	1186	175	126	-44.0
RdDv	2887	1902	3335	1167	206	141	-41.0
RdDd	3029	1520	2997	1042	71	125	-41.2

^atotal exposed surface area

^bnonpolar (carbon) exposed surface area

^cburied surface area for carbon atom (C), uncharged nitrogen/oxygen atoms (N/O), charged oxygen/nitrogen atom (O⁻/N⁺) and sulfur atoms (S)

^dcalculated contribution of hydrophobic energies to free energy of folding (eqn. 2)

Thermodynamic origins of hyperthermostability

Although there have been no detailed calorimetric studies of the stability of proteins from hyperthermophilic organisms and it is not clear to what degree these proteins are more stable than their mesophilic counterparts and understanding of the thermodynamics governing their stability is crucial to an analysis of their stability. Some insights may be derived from studies done on small mesophilic, water soluble proteins. The thermal unfolding of these proteins are generally characterized by assuming a transition between two states, the native (N) and the unfolded (U) form, where the free energy of this transition, ΔG_{NU} , is approximately 10-15 kcal/mol (Privalov, 1979; Privalov and Gill, 1988; Privalov, 1989). This transition is accompanied by a large, positive change in the heat capacity as buried hydrophobic groups become exposed to the solvent.

For a two state process, assuming ΔC_p is constant and independent of temperature, ΔG_{NU} is:

$$\Delta G_{NU} = \Delta H_m \left[\frac{T_m - T}{T_m} \right] - \int_T^{T_m} \Delta C_p dT + T \int_T^{T_m} \Delta C_p d \ln T \quad (3)$$

where T_m is the transition temperature for unfolding during heat denaturation (the temperature at which $\Delta G_{NU} = 0$); and ΔH_m is the enthalpy change at T_m (Privalov, 1979). The entropy change at T_m is given by $\Delta S_m = \Delta H_m/T_m$. Therefore ΔG_{NU} is completely specified by the three parameters, T_m , ΔH_m and ΔC_p , or equivalently, ΔH_m , ΔS_m , and ΔC_p . The dependence of ΔG_{NU} on T for a given set of solution conditions (pH, ionic strength, etc.) defines the stability curve for a protein (Becktel and Schellman, 1987).

Hyperthermostability can in principle be achieved by some combination of increasing ΔH_m and/or decreasing ΔS_m and/or ΔC_p . There are few *a priori* restrictions on how these changes are accomplished, provided that $\Delta H_m/\Delta S_m = T_m > \sim 370$ K. Possible mechanisms for enhancing the thermostability of proteins by alterations in ΔH_m , ΔS_m and ΔC_p will be briefly discussed. The emphasis in this discussion is on interactions that

stabilize the native state, since these can in principle be characterized from the native structure. It is also possible, however, that protein stability reflects contributions from unfolded forms (Dill and Shortle, 1971) that cannot be addressed in this study.

The enthalpy of unfolding may be increased relative to mesophilic rubredoxins for the rubredoxin from *P. furiosus* due to a more extensive network of hydrogen bonds than is found in the mesophilic counterparts. Another factor that may exert a strong influence on the stability of the native state in rubredoxin is the iron ligation, but, considering the similarity, in this respect, to all of the other rubredoxins it is not likely that iron ligation is responsible for the relative differences in thermal stability.

Privalov has shown that ΔC_p is proportional to the number of hydrophobic contacts within the protein (Privalov, 1979) or stated differently, is proportional to the surface area of buried nonpolar residues (Livingstone, Spolar et al., 1991). Recall from above that the amount of nonpolar surface area buried by folding is approximately the same for all of the rubredoxins. Therefore, it is not likely that changes in ΔC_p makes a significant contribution to the increased thermal stability.

Two sources generally contribute to the entropy of unfolding (Privalov, 1979): (1) the change in the number of conformational degrees of freedom upon unfolding, and (2) hydration effects resulting from the exposure of buried sidechains to water. As discussed above it does not appear that hydration effects are vastly different between the rubredoxins so they probably do not make a significant contribution to any change in the entropy of unfolding. However, it is possible that substitutions for glycine and proline in this rubredoxin do have an effect (Nemethy, Leach et al., 1966; Hecht, Sturtevant et al., 1986; Matthews, Nicholson et al., 1987). Substitution of Glu-14 for Pro in RdPf would be expect to increase the entropy of unfolding but this effect could be offset by favorable electrostatic interactions between Glu-14 and surrounding residues. The presence of

Pro-44 in RdPf, usually Gly in other rubredoxins, could stabilize the native state by decreasing the number of degrees of freedom in the unfolded state.

While there are a few more hydrogen bonds and somewhat more nonpolar surface area estimated to be buried upon unfolding in RdPf than other rubredoxins, there is no evidence for large differences in the number of stabilizing interactions, or for new types of interactions that are responsible for RdPf stability. This suggests that it is unlikely that there are enormous differences in stabilization energies for RdPf relative to mesophilic rubredoxins.

References

- Ahern, T. J. and Klibanov, A. M. (1985). "The Mechanism of Irreversible Enzyme Inactivation at 100°C." *Science* **228**, 1280-1283.
- Argos, P., Rossmann, M. G., Grau, U. M., Suborn, H., Frank, G. and Tratschin, J. D. (1979). "Thermal Stability and Protein Structure." *Biochem.* **18**, 5698-5703.
- Becktel, W. J. and Schellman, J. A. (1987). "Protein Stability Curves." *Biopolymers* **26**, 1859-1877.
- Blake, P. R., Day, M. W., Hsu, B. T., Joshua-Tor, L., Park, J. B., Hare, D. R., Adams, M. W. W., Rees, D. C. and Summers, M. F. (1992). "Comparison of the x-ray structure of native rubredoxin from *pyrococcus-furiosus* with the nmr structure of the zinc-substituted protein." *Protein Science* **1**, 1522-1525.
- Blake, P. R., Park, J.-B., Bryant, F. O., Aono, S., Magnuson, J. K., Eccleston, E., Howard, J. B., Summers, M. F. and Adams, M. W. W. (1991). "Determinants of Protein Hyperthermostability: Purification and Amino Acid Sequence of Rubredoxin from the Hyperthermophilic Archaeobacterium *Pyrococcus furiosus* and Secondary Structure of the Zinc Adduct by NMR." *Biochemistry* **30**, 10885-10895.
- Blake, P. R., Park, J.-B., Zhou, Z. H., Hare, D. R., Adams, M. W. W. and Summers, M. F. (1992). "Solution-state Structure by NMR of Zinc-Substituted Rubredoxin From the Marine Hyperthermophilic Archaeobacterium *Pyrococcus furiosus*." *Prot. Sci.* **1**, 1508-1521.
- Bradley, E. A., Stewart, D. E., Adams, M. W. W. and Wampler, J. E. (1993). "Investigations of the thermostability of rubredoxin models using molecular-dynamics simulations." *Protein Science* **2**, 650-665.
- Burley, S. K. and Petsko, G. A. (1985). "Aromatic-Aromatic Interaction: A Mechanism of Protein Structure Stabilization." *Science* **229**, 23-28.
- Chothia, C. (1976). "The nature of the accessible and buried surfaces in proteins." *Nature* **105**, 1-14.
- Chothia, C. and Lesk, A. M. (1986). "The relation between the divergence of sequence and structure in proteins." *The EMBO Journal* **5**, 823-826.
- Davies, G. J., Gamblin, S. J., Littlechild, J. A. and Watson, H. C. (1993). "The Structure of a Thermally Stable 3-Phosphoglycerate Kinase and a Comparison With Its Mesophilic Equivalent." *Proteins: Struct. Func. Genet.* **15**, 283-289.

- Day, M. W., Hsu, B. T., Joshua-Tor, L., Park, J. B., Zhou, Z. H., Adams, M. W. W. and Rees, D. C. (1992). "X-ray crystal-structures of the oxidized and reduced forms of the rubredoxin from the marine hyperthermophilic archaebacterium *pyrococcus-furiosus*." *Protein Science* **1**, 1494-1507.
- Dill, K. and Shortle, D. (1971). "Denatured states of Proteins." *Annu. Rev. Biochem.* **60**, 795-825.
- Eisenberg, D. and McLachlan, A. D. (1986). "Solvation energy in protein folding and binding." *Nature* **319**, 199-203.
- Eisenberg, D., Wesson, M. and Yamashita, M. (1989). "Interpretation of protein folding and binding with atomic solvation parameters." *Chem. Scrip.* **29A**, 217-221.
- Hamann, S. D. (1980). "The Role of Electrostriction in High Pressure Chemistry." *Rev. Phys. Chem. Jpn.* **50**, 147-168.
- Hecht, M. H., Sturtevant, J. M. and Sauer, R. T. (1986). "Stabilization of λ Repressor Against Thermal Denaturation by Site-Directed Gly \rightarrow Ala Changes in α -Helix 3." *Proteins* **1**, 43-46.
- Hei, D. J. and Clark, D. S. (1994). "Pressure stabilization of proteins from extreme thermophiles." *Applied And Environmental Microbiology* **60**, 932-939.
- Heremans, K. (1980). "Biophysical Chemistry at High Pressure." *Rev. Phys. Chem. Jpn.* **50**, 256-273.
- Hvidt, A. (1975). "A Discussion of Pressure-Volume Effects in Aqueous Protein Solutions." *J. Theor. Biol.* **50**, 245-252.
- Imada, K., Sato, M., Tanaka, N., Katsube, Y., Matsuura, Y. and Oshima, T. (1991). "Three-dimensional Structure of a Highly Thermostable Enzyme, 3-Isopropylmalate Dehydrogenase of *Thermus thermophilus* at 2.2Å Resolution." *J. Mol. Biol.* **222**, 725-738.
- Kauzmann, W. (1959). "Some Factors in the Interpretation of Protein Denaturation." *Adv. Prot. Chem* **14**, 1-63.
- Kelly, C. A., Nishiyama, M., Ohnishi, Y., Beppu, T. and Birktoft, J. J. (1993). "Determinants of Protein Thermostability Observed in the 1.9Å Crystal Structure of Malate Dehydrogenase from the Thermophilic Bacterium *Thermus flavus*." *Biochem.* **32**, 3913-3922.
- Klump, H. H., Adams, M. W. W. and Robb, F. T. (1994). "Life in the pressure cooker: The thermal unfolding of proteins from hyperthermophiles." *Pure & Appl. Chem.* **66**, 485-489.
- Lee, B. and Richards, F. M. (1971). "The interpretation of protein structures: Estimation of static accessibility." *J. Mol. Biol.* **55**, 379-400.
- Livingstone, J. R., Spolar, R. S. and Record, M. T. (1991). "Contributions to the thermodynamics of protein folding from the reduction in water-accessible nonpolar surface area." *Biochem.* **30**, 4237-4244.
- Matthews, B. W., Nicholson, H. and Becktel, W. J. (1987). "Enhanced Protein Thermostability From Site-Directed Mutations That Decrease The Entropy of Unfolding." *Proc. Natl. Acad. Sci. USA* **84**, 6663-6667.
- Miller, S., Janin, J., Lesk, A. M. and Chothia, C. (1987). "Interior and surface of monomeric proteins." *J. Mol. Biol.* **196**, 641-656.
- Nemethy, G., Leach, S. J. and Scheraga, H. A. (1966). "The influence of amino acid side chains on the free energy of helix-coil transitions." *Journal of Physical Chemistry* **70**, 998-1004.

- Nicholson, H., Bechtel, W. J. and Matthews, B. W. (1988). "Enhanced Protein Thermostability From Designed Mutations That Interact with α -helix Dipoles." *Nature* **336**, 651-656.
- Perutz, M. F. (1978). "Electrostatic effects in proteins." *Science* **201**, 1187-1191.
- Perutz, M. F. and Raidt, H. (1975). "Stereochemical basis of heat stability in bacterial ferredoxins and in haemoglobin A2." *Nature* **255**, 256-259.
- Privalov, P. L. (1979). "Stability of Proteins: Small Globular Proteins." *Advances in Protein Chemistry* **33**, 167-241.
- Privalov, P. L. (1989). "Thermodynamic problems of protein structure." *Annu. Rev. Biophys. Chem.* **18**, 47-69.
- Privalov, P. L. and Gill, S. J. (1988). "Stability of Protein Structure and Hydrophobic Interaction." *Advances in Protein Chemistry* **39**, 191-234.
- Privalov, P. L. and Khechinashvili, N. N. (1974). "A Thermodynamic Approach to the Problem of Stabilization of Globular Protein Structure: A Calorimetric Study." *J. Mol. Biol.* **86**, 665-684.
- Richards, F. M. (1977). "Areas, volumes, packing and protein structure." *Ann. Rev. Biophys. Bioeng.* **6**, 151-276.
- Skarzynski, T., Moody, P. C. E. and Wonacott, A. J. (1987). "Structure of holo-Glyceraldehyde-3-phosphate Dehydrogenase from *Bacillus stearothermophilus* at 1.8Å Resolution." *J. Mol. Biol.* **193**, 171-187.
- Spezio, M., Wilson, D. B. and Karplus, P. A. (1993). "Crystal Structure of the Catalytic Domain of a Thermophilic Endocellulase." *Biochem.* **32**, 9906-9916.
- Walker, J. E., Wonacott, A. J. and Harris, J. I. (1980). "Heat Stability of a Tetrameric enzyme, D-glyceraldehyde-3-phosphate dehydrogenase." *Eur. J. Biochem.* **108**, 581-586.

Chapter 2
Nitrogenase MoFe Protein
from
Azotobacter vinelandii

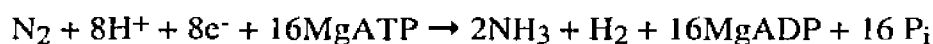
2.1 Introduction to Nitrogenase MoFe

There are essentially three global reservoirs for nitrogen – the atmosphere, the oceans and the terrestrial ecosystems. Over 99.9% of the nitrogen in these global reservoirs is inaccessible to nearly all living organisms (Kinzig and Socolow, 1994) although nitrogen in its reduced form is an absolutely essential element for all living systems. This large reservoir of nitrogen is diatomic nitrogen, the N_2 gas present in the atmosphere or dissolved in the oceans. Nitrogen is the most abundant gas in the atmosphere and accounts for nearly 80% of the total gas content. The strength of the $N\equiv N$ triple bond renders diatomic nitrogen extremely stable to the majority of conditions likely to be found in living systems. Nonetheless, reduced nitrogen is an essential component of most molecules of any biological significance, including proteins and nucleic acids. The question then arises – How does nitrogen get transformed from its thermodynamically stable or kinetically inert form (N_2) into a biologically useful and accessible form (generally NH_3)?

Although the reduction of dinitrogen to ammonia is a thermodynamically favorable reaction, a tremendous amount of energy must be supplied to activate the process. Lightning in the atmosphere accounts for the major natural non-biological source of reduced nitrogen, but the action of lightning (even at approximately 3 billion flashes per year) amounts to only a very small part of the total flow of usable nitrogen into the biosphere. The dinitrogen processed by lightning flashes is first oxidized to nitric oxide (NO) and then converted to NO_3^- by rain. Another energetic process is used by industry to reduce dinitrogen. This is the Haber-Bosch process which requires an Fe-catalyst and temperatures between 300 and 500° C, along with dinitrogen and dihydrogen (H_2) pressures in excess of 300 atmospheres, to produce ammonia. The Haber-Bosch process amounts for approximately 25% of the total influx of reduced nitrogen into the biosphere (Burns and Hardy, 1975). In contrast there are a small number of microorganisms that

account for nearly 60% of the total conversion of dinitrogen into a biologically useful form. These organisms are termed diazotrophs and accomplish the transformation under extremely mild conditions: pressures of 0.8 atmospheres of N₂ and temperatures of 20 to 25° C. These are, of course, ambient conditions and it would be safe to say that all living organisms are in some way dependent on the diazotrophs for their supply of nitrogen.

The processes mentioned above are termed nitrogen fixation. The biological fixation of nitrogen is accomplished through an enzymatic process. The overall stoichiometry of this process is



The protein system that carries on this process is nitrogenase. The nitrogenase system has two major components, the iron protein and the molybdenum-iron protein.

The Fe-protein is a homodimer with a total molecular mass of ~60 kDa. Each monomer folds into a single α -helical/ β -sheet domain and the two monomers ligand a single 4Fe:4S cubane at one end of their interface. The electrons required for the reduction process are transferred from the Fe-protein, coupled to the hydrolysis of ATP, to the molybdenum-iron protein (MoFe), the actual site of dinitrogen reduction.

The MoFe-protein is a tetrameric $\alpha_2\beta_2$ unit composed of a dimer of $\alpha\beta$ dimers. The α and β -subunits are coded by the *nifD* and *nifK* genes respectively. Each of the four subunits contain three domains that adopt an α -helical/ β -sheet type of fold. The α -subunit of the MoFe protein from *Azotobacter vinelandii* contains 491 amino acids and the β -subunit contains 522. The total molecular mass of the tetramer (the minimal functional unit of MoFe) is approximately 200 kD. The α -subunit contains the molybdenum-iron cofactor (M-cluster, shown in Figure 2.3-8) which is believed to be the actual site of dinitrogen reduction (Hawkes, McLean, 1984). The structure of the M-cluster was unprecedented. The other metal center of MoFe, the P-cluster pair (shown in Figure 2.3-10), was also an unprecedented type of double cubane composed of two 4Fe:4S

cubanes bridged by two thiols and containing a disulfide bond. The P-cluster pair bridges both the α -subunit and the β -subunit at an approximate two-fold rotation axis between the two homologous sub-units. The P-cluster pair is believed to pass the electrons generated by the Fe-protein to the M-cluster.

This chapter (Chapter 2) will discuss the high resolution (2.2Å resolution) refinement of the crystal structure of the MoFe protein from *Azotobacter vinelandii* as determined by Kim and Rees (Kim and Rees, 1992; Kim and Rees, 1992; Kim and Rees, 1993) (section 2.2) and will describe the structural features of the molecule (section 2.3) followed by a discussion of these features with respect to their functional significance (section 2.4). Several excellent reviews have been written on the structure and function of nitrogenase (Orme-Johnson, 1985; Burgess, 1990; Eady, 1991; Smith and Eady, 1992; Burgess, 1993; Newton and Dean, 1993; Orme-Johnson, 1993; Rees, 1993; Eady and Leigh, 1994; Howard and Rees, 1994; Kim and Rees, 1994; Leigh, 1994). The reader is referred to these for an in-depth review of the topic. The discussion that follows in this section (section 2.1) is a brief overview of what was known before the structure was solved and what insights the structure has had on our understanding of the function of nitrogenase.

The Metal Centers

The structure of the metal centers of the MoFe-protein is an area that has attracted much interest and debate. Before the solution of the crystal structure, much was known about the M-cluster based on spectroscopic investigations. Early studies had shown that the ratio of iron to molybdenum in the FeMo-cofactor was 6-7Fe:1Mo (Yang, Pan, 1982). Other investigators demonstrated that the ratio of sulfur to molybdenum was 8-9S:1Mo (Nelson, Levy, 1983; Smith, Bishop, 1985). These results taken with the results of Hoover et al. as to the requirement for homocitrate (Hoover, Robertson, 1987) led to a

consensus composition for the cofactor of 1Mo:6-7Fe:8-9S:1homocitrate. Information as to the general organization of these atoms with respect to each other was obtained through the use of Mo K absorption edge EXAFS. The first EXAFS studies revealed the cluster nature of the FeMo-cofactor by indicating that iron and sulfur were nearest neighbors at distances consistent with sulfide-like bridging distances (Cramer, Gillum, 1978; Cramer, Hodgson, 1978). Later studies indicated that soft ligands like oxygen and/or nitrogen were present (Conradson, Burgess, 1987). The results obtained by Cramer et al. were limited by the scarcity of model compounds on which to base the analysis, and the results obtained by Conradson et al. profited greatly by the abundance of model compounds spawned by the information provided by Cramer. The determination of the actual structure of the FeMo-cofactor has allowed a more comprehensive analysis of the EXAFS results and has revealed a feature of the spectra that is consistent with the second set Mo to Fe distances of approximately 5Å seen in the crystal structure (Liu, Filipponi, 1994).

Many model compounds had been synthesized that imitated the spectroscopic signature of the FeMo-cofactor and/or reduced acetylene or other non dinitrogen substrates of nitrogenase (Coucouvani, 1991). These synthetic studies were conducted with the goal of finding an inorganic chemistry solution to what was once thought to be an insurmountable biochemical and crystallographic problem, i.e. the mechanism of dinitrogen reduction and the structure of the species that were responsible. The subsequent elucidation of the structure for the metal centers has been met with mixed emotions; however, the model of the M-cluster has been generally accepted as correct. Nevertheless, the precise composition of the P-cluster pair is an area that is still open for discussion, and evidence will be presented in section 2.2 that supports the model of Kim and Rees. As of this time, there have been no reports of the successful synthesis of either the M-cluster or the P-cluster pair although there have been a number of interesting compounds reported that share either structural and/or functional characteristics with the M-cluster (Cen,

MacDonnell, 1994; Demadis and Coucouvanis, 1995; Laughlin and Coucouvanis, 1995; Malinak, Demadis, 1995).

The polypeptide environment around the FeMo-cofactor contains a number of residues that have been targeted for mutagenesis studies. For example, Cys α -275 and His α -442 have been identified as ligands to the M-cluster and any mutation at either of these two residues results in an inactive protein unable to reduce nitrogen (Brigle, Setterquist, 1987; Kent, Ioannidis, 1989). The ligation of the cofactor by Cys α -275 had been proposed on the basis of sequence analysis, and was confirmed by the crystal structure, but the discovery of His α -442 as a ligand was a surprise. Other surprises arising from the crystal structure were the proximity of Arg α -96, Phe α -381, Arg α -359 and Gln α -440 to the FeMo-cofactor.

The two residues Gln α -191 and His α -195 were identified as being potentially important residues in the FeMo-cofactor environment because they were two residues that are highly conserved in nitrogenase and were not present in NifE, a protein homologous to the α -subunit that participates in the biosynthesis of the M-cluster (Scott, May, 1990). In NifE, where a precursor to the FeMo-cofactor is believed to be synthesized, prior to transfer to the *NifD*-gene product, these residues are Lys and Asn respectively, which led to the prediction that they might be cofactor ligands.

When Gln α -191 was replaced with Lys, the result was a protein unable to reduce N_2 (i.e. was *Nif*⁻) but able to still reduce protons and acetylene and was sensitive to CO (Scott, Dean, 1992). Scott et al. also demonstrated that this change in activity was a result of a change in the environment around the FeMo-cofactor and not a change in the cofactor itself. Another interesting consequence of this mutation was that now the MoFe protein was able to reduce acetylene by either two or four electrons to either ethylene or ethane respectively, while the native form of MoFe is only able to catalyze the two electron process (Dilworth, 1966). Similar results were obtained when His α -195 was mutated to

Asn (Scott, Dean, 1992). It is interesting to note that the four electron reduction of acetylene to ethane is a process that the vanadium-dependent nitrogenase also catalyzes, but presumably by a different mechanism (Dilworth, Eady, 1988; Scott, Dean, 1992). Not all mutations to Gln α -191 produced *Nif*⁻ mutants, however the Ser mutant grows slowly and the Ala and Pro mutants have been shown to exhibit diazotrophic growth rates comparable to wild type (Newton and Dean, 1993). In contrast, all mutations of His α -195 appear to be *Nif*⁻.

Electron spin echo envelope modulation (ESEEM) experiments had suggested that the FeMo-cofactor was liganded by a nitrogen, probably a His (Thomann, Morgan, 1987). A feature of the His α -195 to Asn mutant is the disappearance of the ESEEM signal concomitant with the loss of N₂ reduction (Thomann, Bernardo, 1991). This observation led to the prediction that His α -195 was a ligand to the cofactor and contributed to the surprise that His α -442 was the ligand. The mutation of other conserved histidines, for example His α -83, His α -196, His α -274 or His β -90, did not produce a loss of the ESEEM signal or stop N₂ reduction. The mutation His α -195-Gln produces no change in the ESEEM signal, however, and results in a phenotype where acetylene and H₂ evolution are competitively inhibited by N₂ (DeRose, Kim, 1995). This result suggests that the ϵ -nitrogen of the imidazole ring and the amino-nitrogen atom of Gln form the same hydrogen bond to a bridging sulfur of the FeMo-cofactor and that this hydrogen bond helps to orient the cofactor in its binding pocket so that His α -442 can generate the ESEEM signal.

Other residues found in the FeMo-cofactor pocket have been, while their specific roles have yet to be precisely defined (Newton and Dean, 1993). Based on the structure, it is likely that these residues may 1) be involved in the transfer of the FeMo-cofactor into the binding site, 2) participate in proton transfer or 3) provide interactions that stabilize the orientation of the cofactor in the polypeptide pocket. Mutation of Gln α -440 to Glu has no

effect on the catalytic activity of MoFe. This is reasonable because the crystal structure shows that Gln α -440 forms a water mediated hydrogen bond to a terminal carboxylate of the homocitrate. Substitution of Gln for Arg α -96 and Arg α -359 results in a slow growth phenotype for the Arg α -96-Gln mutant and a phenotype that cannot grow diazotrophically for the Arg α -359-Gln mutant. When Phe α -381 is substituted by Arg a *Nif⁻* phenotype results. The close approach of Phe α -381 to two bridging sulfurs of the FeMo-cofactor may not allow the additional bulk of an Arg without significant disruption of the cofactor orientation.

References

- Brigle, K. E., Setterquist, R. A., Cantwell, D. R. D. J. S., Weiss, M. C. and Newton, W. E. (1987). "Site-Directed Mutagenesis of the Nitrogenase MoFe protein of *Azotobacter vinelandii*." *Proc. Natl. Acad. Sci.* **84**, 7066-7069.
- Burgess, B. K. (1990). "The Iron-Molybdenum Cofactor of Nitrogenase." *Chem. Rev.* **90**, 1377-1406.
- Burgess, B. K. (1993). "Nitrogenase Structure, Function, and Genetics." *ACS Symp. Ser.* **535**, 144-169.
- Burns, R. C. and Hardy, R. W. F. (1975). Nitrogen Fixation in Bacteria and Higher Plants. In (Editor), Springer-Verlag, Berlin.
- Cen, W., MacDonnell, F. M., Scott, M. J. and Holm, R. H. (1994). "Heterometal Clusters Containing the Cuboidal Fe_4S_3 Fragment: Synthesis, Electron Distribution, and Reactions." *Inorg. Chem.* **33**, 5809-5818.
- Conradson, S. D., Burgess, B. K., Newton, W. E., Mortenson, L. E. and Hodgson, K. O. (1987). "Structural Studies of the Molybdenum Site in the MoFe Protein and Its FeMo Cofactor by EXAFS." *J. Am. Chem. Soc.* **109**, 5364.
- Coucouvanis, D. (1991). "Use of Preassembled Fe/S and Fe/Mo/S Clusters in the Stepwise Synthesis of Potential Analogues for the Fe/Mo/S Site in Nitrogenase." *Acc. Chem. Res.* **24**, 1-8.
- Cramer, S. P., Gillum, W. O., Hodgson, K. O., Mortenson, L. E., Stiefel, E. I., Chisnell, J. R., Brill, W. J. and Shah, V. K. (1978). *J. Am. Chem. Soc.* **100**, 3814.
- Cramer, S. P., Hodgson, K. O., Gillum, W. O. and Mortenson, L. E. (1978). *J. Am. Chem. Soc.* **100**, 3398.
- Demadis, K. D. and Coucouvanis, D. (1995). "Synthesis, Structural Characterization, and Properties of New Single and Double Cubanes Containing the MoFe_3S_4 Structural Unit and Molybdenum-Bound Polycarboxylate Ligands. Clusters with a Molybdenum-Coordination Environment Similar to That in the Iron-Molybdenum Cofactor of Nitrogenase." *Inorg. Chem.* **34**, 436-448.
- DeRose, V. J., Kim, C. H., Newton, W. E., Dean, D. R. and Hoffman, B. M. (1995). "Electron Spin Echo Envelope Modulation Analysis of Altered Nitrogenase MoFe Proteins from *Azotobacter vinelandii*." *Biochem.* **34**, 2809-2814.

- Dilworth, M. J. (1966). "Acetylene Reduction by Nitrogen-Fixing Preparations from *Clostridium pasteurianum*." *Biochem. Biophys. Acta.* **127**, 285-294.
- Dilworth, M. J., Eady, R. R. and Eldridge, M. E. (1988). "The Vanadium Nitrogenase of *Azotobacter chroococcum* - Reduction of Acetylene and Ethylene to Ethane." *Biochem. J.* **249**, 745-751.
- Eady, R. R. (1991). "The Mo-, V-, Fe-Based Nitrogenase Systems of *Azotobacter*." *Adv. Inorg. Chem.* **36**, 77-102.
- Eady, R. R. and Leigh, G. J. (1994). "Metals in the Nitrogenases." *J. Chem. Soc. Dalton Trans.* 2739-2747.
- Hawkes, T. R., McLean, P. A. and Smith, B. E. (1984). "Nitrogenase From *nifV* Mutants of *Klebsiella pneumoniae* Contains an Altered Form of The Iron Molybdenum Cofactor." *Biochem.* **217**, 317-321.
- Hoover, T. R., Robertson, A. D., Cerny, R. L., Hayes, R. N., Imperial, J., Shah, V. K. and Ludden, P. W. (1987). "Identification of the V-Factor Needed for Synthesis of the Iron-Molybdenum Cofactor of Nitrogenase as Homocitrate." *Nature* **329**, 855-857.
- Howard, J. B. and Rees, D. C. (1994). "NITROGENASE: A Nucleotide-Dependent Molecular Switch." *Annu. Rev. Biochem.* **63**, 235-264.
- Kent, H. M., Ioannidis, I., Gormal, C., Smith, B. E. and Buck, M. (1989). "Site-directed Mutagenesis of *Klebsiella pneumoniae* Nitrogenase." *Biochem. J.* **264**, 257-264.
- Kim, J. and Rees, D. C. (1992). "Crystallographic structure and functional implications of the nitrogenase molybdenum-iron protein from *Azotobacter vinelandii*." *Nature* **360**, 553-560.
- Kim, J. and Rees, D. C. (1992). "Structural Models for the Metal Center in the Nitrogenase Molybdenum-Iron Protein." *Science* **257**, 1677-1682.
- Kim, J. and Rees, D. C. (1993). "X-ray Crystal Structure of the Nitrogenase Molybdenum-Iron Protein from *Clostridium pasteurianum* at 3.0 Å Resolution." *Biochemistry* **32**, 7104-7115.
- Kim, J. and Rees, D. C. (1994). "Nitrogenase and Biological Nitrogen Fixation." *Biochemistry* **33**, 387-397.
- Kinzig, A. P. and Socolow, R. H. (1994). "Human Impacts on the Nitrogen Cycle." *Physics Today* November, 24-31.
- Laughlin, L. J. and Coucouvanis, D. (1995). "Use of $[\text{MoFe}_3\text{S}_4]^{3+}$ Single Cubanes in the Catalytic Reduction of Acetylene to Ethylene and Ethane. Identification of Molybdenum and Iron Atoms as Catalytic Sites during Substrate Reduction and Implications for Nitrogenase Action." *J. Am. Chem. Soc.* **117**, 3118-3125.
- Leigh, G. J. (1994). "Dinitrogen coordination chemistry and the new nitrogenases." *New J. Chem.* **18**, 157-161.
- Liu, H. I., Filipponi, A., Gavini, N., Burgess, B. K., Hedman, B., Di Cicco, A., Natoli, C. R. and Hodgson, K. O. (1994). "EXAFS Studies of FeMo-Cofactor and MoFe Protein: Direct Evidence for the Long-Range Mo-Fe-Fe Interaction and Cyanide Binding to the Mo in FeMo-Cofactor." *J. Am. Chem. Soc.* **116**, 2418-2423.

- Malinak, S. M., Demadis, K. D. and Coucouvanis, D. (1995). "Catalytic Reduction of Hydrazine to Ammonia by the VF3S4 Cubanes. Further Evidence for the Direct Involvement of the Heterometal in the Reduction of Nitrogenase Substrates and Possible Relevance to the Vanadium Nitrogenases." *J. Am. Chem. Soc.* **117**, 3126-3133.
- Nelson, M. J., Levy, M. A. and Orme-Johnson, W. H. (1983). "Metal and Sulfur Composition of Iron Molybdenum Cofactor of Nitrogenase." *Proc. Natl. Acad. Sci.* **260**,
- Newton, W. E. and Dean, D. R. (1993). "Role of the Iron-Molybdenum Cofactor Polypeptide Environment in *Azotobacter vinelandii* Molybdenum-Nitrogenase Catalysis." *ACS Symp. Ser.* **535**, 216-230.
- Orme-Johnson, W. H. (1985). "Molecular Basis of Biological Nitrogen Fixation." *Ann. Rev. Biophys. Biophys. Chem.* **14**, 419-459.
- Orme-Johnson, W. H. (1993). "The Molybdenum-Iron Protein of Nitrogenase: Structural and Functional Feature of Metal Cluster Prosthetic Groups." *ACS Symp. Ser.* **535**, 257-270.
- Rees, D. C. (1993). "Dinitrogen reduction by nitrogenase: if N_2 isn't broken, it can't be fixed." *Curr. Opin. Struc. Biol.* **3**, 921-928.
- Scott, D. J., Dean, D. R. and Newton, W. E. (1992). "Nitrogenase-catalysed Ethane Production and CO-sensitive Hydrogen Evolution from MoFe Proteins Having Amino Acid Substitutions in an α -Subunit FeMo Cofactor-binding Domain." *J. Biol. Chem.* **267**, 20002-20010.
- Scott, D. J., May, H. D., Newton, W. E., Brigle, K. E. and Dean, D. R. (1990). "Role for the Nitrogenase MoFe protein α -subunit in the FeMo-cofactor binding domain." *J. Biol. Chem.* **267**, 20002-20010.
- Smith, B. E., Bishop, P. E., Dixon, R. A., Eady, R. R., Filler, W. A., Lowe, D. J., Richards, A. J. M., Thompson, A. J., Thornely, R. N. F. and Postgate, J. R. (1985). Nitrogen Fixation Research Progress. Martinus Nijhoff. Dordrecht, The Netherlands; pp. 597.
- Smith, B. E. and Eady, R. R. (1992). "Metalloclusters of the nitrogenases." *Eur. J. Biochem.* **205**, 1-15.
- Thomann, H., Bernardo, M., Newton, W. E. and Dean, D. R. (1991). "N Coordination of the FeMo-cofactor requires His-195 of The MoFe Protein α Subunit and is Essential For Biological Nitrogen Fixation." *Proc. Natl. Acad. Sci.* **88**, 6620-6623.
- Thomann, H., Morgan, T. V., Jin, H., Burgmayer, S. J. N., Bare, R. E. and Stiefel, E. I. (1987). "Protein Nitrogen Coordination to the FeMo center of Nitrogenase from *Clostridium pasteurianum*." *J. Am. Chem. Soc.* **109**, 7913-7914.
- Yang, S. S., Pan, W. H., Friesen, G. D., Burgess, B. K., Corbin, J. L., Steifel, R. I. and Newton, W. E. (1982). *J. Biol. Chem.* **257**, 8042.

2.2 Experimental

2.2.1 Structure Solution

Although the work described here in Chapter 2 pertaining to the MoFe nitrogenase from *Azotobacter vinelandii* (Av1) involves the **refinement only**, a brief description of the crystallization, collection and processing of the diffraction data, and solution of the structure, are included here for the convenience of the reader. For more complete details the reader should refer to the references cited herein.

Crystallization

Crystals of the MoFe-protein from *Azotobacter vinelandii* (Av1) were grown by the microcapillary batch method as described by Kim and Rees (Kim and Rees, 1992). Monoclinic crystals grow from 18% polyethelene glycol (PEG) 4000, 0.14 M NaCl, 0.10 M Na₂MoO₄, 80 mM tris-HCl, pH 8.0 and ~8mg/ml protein with $a=108.4\text{\AA}$, $b=130.5\text{\AA}$, $c=81.5\text{\AA}$ and $\beta=110.8^\circ$ in space group P 2₁. Data collected from this crystal form was used for the refinement discussed in the following section (2.2.2).

Two crystal forms of the MoFe-protein from *Clostritium pasteurianum* were used for electron density averaging in the determination of the structure of MoFe from *A. vinelandii*. Monoclinic crystals (space group P 2₁) with $a=70.0\text{\AA}$, $b=151.3\text{\AA}$, $c=121.9\text{\AA}$ and $\beta=110.4^\circ$ grow from 15% PEG 4000, 0.21M MgCl, 80mM tris HCl pH .80 and ~8 mg/ml protein. Another crystal form grew in the same space group from 18% PEG 4000, 0.3M CsCl and 80mM tris HCl pH 8.0 (~8 mg/ml protein) with $a=87.9\text{\AA}$, $b=171.4\text{\AA}$, $c=73.6\text{\AA}$ and $\beta=91.5^\circ$.

Data Collection

The data used to solve the structure of Av1 was collected as previously described by Kim and Rees (Kim and Rees, 1992). In summary, native x-ray diffraction data was collected from two crystals of Av1 to a limiting resolution of 2.75\AA and diffraction data

from four heavy atom derivatives (EMTS, PTCL and PIP) for use in MIR phasing were collected to a limiting resolution of 3.0Å. All diffraction data were collected at room temperature with a Siemens multiwire area detector and processed with XENGEN (Howard, Gilliland et al., 1987). The native data set was 94% complete with a merging residual between symmetry-related reflections of 8.5%. The derivative data sets ranged from between 83 to 92% complete with merging residuals between 5.7 to 11.%. The phasing power of the derivatives ranged from a low of 1.17 for PIP and 1.79 for EMTS.

The high resolution data were collected and processed as previously described by Chan et al. (Chan, Kim et al., 1993) and proceeded as follows. X-ray diffraction data to a limiting resolution of 2.2Å were collected from two crystals of Av1 at the Stanford Synchrotron Radiation Laboratory (SSRL) using a MAR Research imaging plate detector. A total of 258,519 observations of 95,078 unique reflections were processed with the MOSFILM and CCP4 packages (SERC, 1986). The final data set was 90% complete to 2.2Å with a merging residual of 13.6%. It is this data set that was used for the structural refinement of Av1 discussed in section 2.2.2.

Structure Solution

The structure was solved by a combination of multiple isomorphous replacement (MIR) and noncrystallographic symmetry (NCS) averaging (Kim and Rees, 1992; Kim and Rees, 1992; Kim and Rees, 1993). A total of three crystal forms (all in the monoclinic space group $P 2_1$) from two organisms (*Azotobacter vinelandii* and *Clostridium pasteurianum*) were required to solve the structure. Heavy atom positions were determined from isomorphous difference patterns. The NCS relationships required for electron density averaging were determined from rotation and translation functions and averaging was performed both within and between crystal forms.

References

- Chan, M. K., Kim, J. and Rees, D. C. (1993). "The Nitrogenase FeMo-Cofactor and P-Cluster Pair: 2.2Å Resolution Structures." *Science* **260**, 792-794.
- Howard, A. J., Gilliland, G. L., Finzel, B. C., Poulos, T. L., Ohlendorf, D. H. and Salemme, F. R. (1987). "The Use of an Imaging Proportional Counter in Macromolecular Crystallography." *Journal of Applied Crystallography* **20**, 383-387.
- Kim, J. and Rees, D. C. (1992). "Crystallographic structure and functional implications of the nitrogenase molybdenum-iron protein from *Azotobacter vinelandii*." *Nature* **360**, 553-560.
- Kim, J. and Rees, D. C. (1992). "Structural Models for the Metal Center in the Nitrogenase Molybdenum-Iron Protein." *Science* **257**, 1677-1682.
- Kim, J. and Rees, D. C. (1993). "X-ray Crystal Structure of the Nitrogenase Molybdenum-Iron Protein from *Clostridium pasteurianum* at 3.0Å Resolution." *Biochemistry* **32**, 7104-7115.
- SERC, D. L. (1986). "CCP4, A Suite of Programs for Protein Crystallography." *SERC Daresbury Laboratory* Warrington, England,

2.2.2 Structure Refinement

The starting coordinates for the high resolution refinement of the MoFe protein from *Azotobacter vinelandii* were those described by Kim and Rees (Kim and Rees, 1992) and listed in the Brookhaven Protein Data Bank (Bernstein, Koetzle et al., 1977; Abola, Bernstein et al., 1987) as 1MIN. These coordinates included 1980/2026 amino-acid residues, two divalent cations (presumably Ca^{+2}), 4 water molecules (liganded to the divalent cations), 2 homocitrate molecules (14 atoms each) and the 66 metal atoms comprising the four cofactors (17 in each MoFe-cofactor and 16 in each P-cluster pair) for a 98% complete model with a total of 15,758 non-hydrogen atoms. The missing amino-acid residues are all part of the α -subunit; $\alpha 2$ - $\alpha 4$, $\alpha 36$ - $\alpha 44$ and $\alpha 482$ - $\alpha 492$. These coordinates were the result of refinement in XPLOR (Brünger, Kuriyan et al., 1987) using a slowcooled simulated annealing protocol.

Two refinement packages were used during the refinement of the model, TNT (Tronrud, Ten Eyck et al., 1987) and XPLOR (Brünger, Kuriyan et al., 1987). The intent was to minimize the introduction of bias in the model due to the refinement package used. Presumably the average of the two resulting coordinates sets would best reflect the true geometry of the metal cofactors. The protein geometry dictionaries of each refinement package were used as supplied and entries for the cofactors were added (see Appendix 2 for the metal cofactor dictionaries that were used). Cofactor entries (distances and angles) were based on small molecule structure determinations where applicable. In both refinement packages all data between 25-2.2Å resolution were used. No data were excluded on the basis of intensity and a sigma cutoff was not employed. In each round of refinement, the positional parameters were refined separately from the temperature factors.

A typical round of refinement in each package included the following steps: (1) Calculation of two electron density maps from the refined coordinates using $F_o - F_c$ (where F_o is the observed structure factor amplitudes and F_c is the calculated structure factor

amplitudes with phases) coefficients for one map and using $2F_o - F_c$ coefficients for the other. (2) The model was then evaluated against the maps resulting from each refinement package using the program TOM/FRODO (Jones, 1978). Amino acid residues were either built to new conformations, removed from the model or added to the model where the electron density indicated it. Solvent atoms were built into the model where it was appropriate (the criteria was good $F_o - F_c$ density, i.e., $>2\sigma$, and reasonable hydrogen bonding geometry). (3) The resulting modified model was then refined.

Table 2.2.2-1 Summary of refinement progress showing the results from each cycle for each refinement package used as well as the total number of atoms and number of solvent oxygens in the model.

Coordinates	XPLOR	TNT	# of Atoms	Solvents
	R-factor	R-factor		
MoFe2	—	22.6%	15758	4
MoFe3	23.8%	22.1%	—	70
MoFe4	22.7%	—	—	155
MoFe5	22.3%	—	15947	205
MoFe6	23.3%	—	16001	287
MoFe7	21.2%	—	16098	384
MoFe8	20.3%	—	16163	540
MoFe9	19.6%	17.5%	16412	617
MoFe10	19.5%	17.3%	16390	625
MoFe11	19.3%	—	16492	643
MoFe12	19.3%	—	16492	640
MoFe13	17.9%	17.4%	16480	625
MoFe14	17.6%	17.9%	16495	625

The protein model (including the cofactors) was never recycled! Each round of refinement began with the protein coordinates reported by Kim and Rees (Kim and Rees, 1992) (1MIN) after they had been manually modified in the program TOM/FRODO according to the electron density maps calculated from the previous round of refinement. The refined solvent shell, however, was always carried over into the next refinement cycle.

In the final rounds of refinement, the very low resolution data was excluded and the data from increasing high resolution shells were included gradually until the entire resolution range of 8.0-2.2Å was included. During the early stages of these refinement cycles, the geometry of the protein was not tightly restrained, and then increasing weights were given to the geometry towards the end. It was assumed that this strategy would allow the data to drive the model away from the current minimum and thereby avoid being trapped at a local minimum. The results would seem to indicate that this was indeed true for the XPLOR refinement. This strategy seemed to have little effect on the refinement with TNT as reflected by the difference in R_{factor} for refinement cycles MoFe9 (where only one resolution range was used) and MoFe13 (See Table 2.2.2-1).

TNT Refinement

Refinement of Av1 with TNT (Tronrud, Ten Eyck et al., 1987) resulted in an $R_{\text{factor}}=17.9\%$. At this stage of refinement the model contains 15,772 protein atoms, 625 solvent oxygen atoms, 2 Ca^{+2} ions and the 94 atoms that comprise the metal cofactors (including the homocitrate). The protein geometry was close to the target values with the rms. deviation from the targets being 0.02Å for the bond distances and 2.7° for the bond angles. Although TNT allows the user to control the weight used for temperature factor correlation between atoms bonded to each other, the temperature factors are not strongly correlated by TNT and they range from the minimum ($B=2.0\text{Å}^2$) to the maximum ($B=100\text{Å}^2$) allowed values. The average B for mainchain atoms is 13.96Å^2 and is 18.65Å^2 for the sidechain atoms with an rms. deviation of the B's for adjacent atoms of 1.61Å^2 . The average solvent B-factor is 24.17Å^2 . The temperature factors for the solvent range from a low of 2.0Å^2 to a high of 74.4Å^2 . The overall average temperature factors for mainchain and sidechain atoms of each subunit are listed in Table 2.2.2-2. A plot of the temperature factors for main chain atoms and for side chain atoms for each subunit is shown in Figure 2.2.2-1 (Compare with Figure 2.2.2-5). As might be expected, the

regions of the polypeptide chain with the largest average temperature factors are regions on the surface of the protein or are solvent exposed regions of the tetramer interface. A Luzzati analysis of the crystallographic residual as a function resolution (Luzzati, 1952) indicates an average coordinate error of approximately 0.15\AA (see Figure 2.2.2-2). Representative electron density for the polypeptide chain is shown in Figure 2.2.2-3. The Ramachandran plot (Srinivasan and Parthasarathy, 1976) of the ϕ , ψ torsion angles is shown in Figure 2.2.2-4.

Table 2.2.2-2. The overall average temperature factors (\AA^2) for main chain and side chain atoms of each subunit of MoFe resulting from refinement in TNT or XPLOR.

	TNT		XPLOR	
	Side Chain	Main Chain	Side Chain	Main Chain
α (A)	19.62	14.86	17.78	15.78
α (C)	20.31	15.30	18.40	16.57
β (B)	17.95	13.21	15.18	13.64
β (D)	16.86	12.48	14.60	12.81

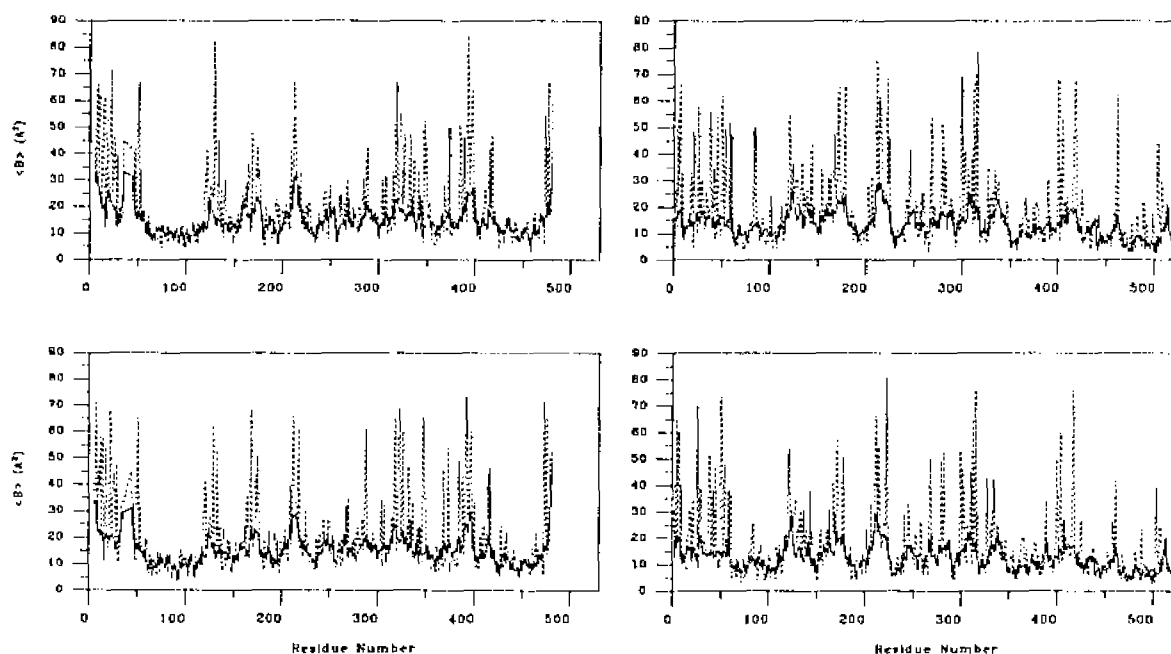


Figure 2.2.2-1. Plot of average temperature factors vs. residue number for MoFe after refinement with TNT. Average side chain B-factors are shown as dotted lines, and average main chain B's are illustrated by the solid lines. The two α -subunits (A and C respectively) are shown in the upper and lower left-hand panels, and the upper and lower right-hand panels represent the two β -subunits (B and D).

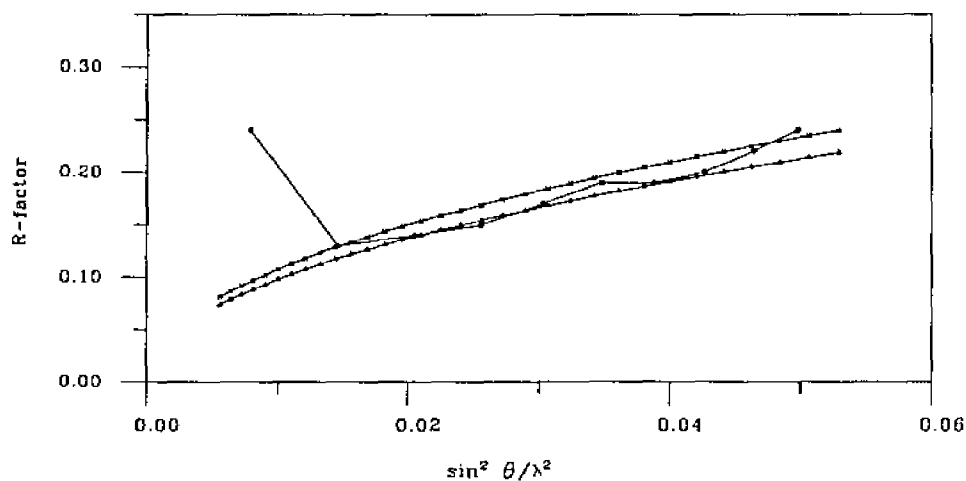


Figure 2.2.2-2. Plot of the R-factor from TNT refinement as a function of the resolution. The upper line corresponds to the theoretical curve for an estimated average coordinate error of 0.22 Å and the lower line is the theoretical curve corresponding to 0.20 Å.

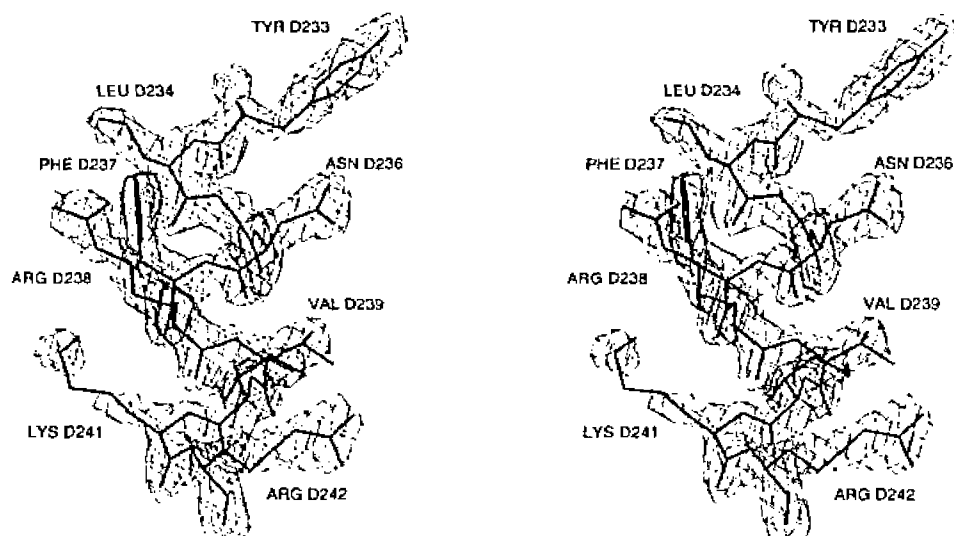


Figure 2.2.2-3. Stereo view of representative electron density of the Mofe protein after refinement with TNT. The density for this helix is from the 2Fo-Fc map and is contoured at the 2σ level. Only residues D233 thru D242 are shown for clarity.

XPLOR Refinement

Refinement in XPLOR (Brünger, Kuriyan et al., 1987) resulted in a crystallographic residual of 17.6% for 16,495 atomic positions. The composition of the model is identical to the model used in the TNT refinement. The rms. deviation in the bond distances and angles are 0.022Å and 2.37° respectively. The default restraints on the temperature factors of adjacent atoms (i.e., 1-2 and 1-3 pairs) was relaxed by one sigma unit. The average temperature factors for mainchain and sidechain atoms are respectively

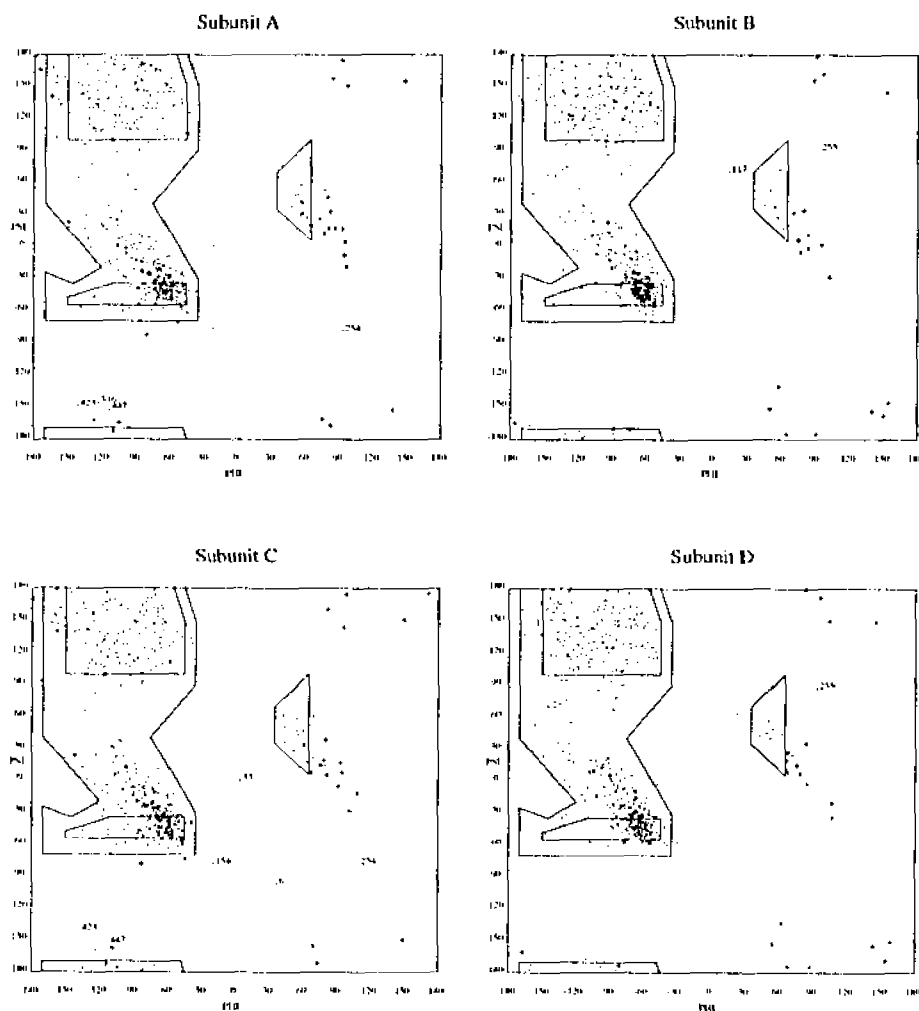


Figure 2.2.2-4. Ramachandran plot of the ϕ , ψ angles for each subunit of MoFe. Glycine residues are represented by + and residues with ϕ , ψ torsion angles outside of the normally allowed regions are labeled by the residue number.

14.70Å² and 16.49Å², ranging from a low of 2.0Å² to a high value of 55.88Å² with an rms. deviation in adjacent B's of 2.67Å² (see Figure 2.2.2-5 and Table 2.2.2-2). Again, here as in the TNT refinement, the largest average temperature factors are observed for regions that are exposed to the solvent. However, unlike in refinement with TNT, the average temperature factors for the side chain atoms are more nearly equal to the average for the main chain atoms of the same residue. The solvent oxygen temperature factor ranges from 2.0 to 57.7Å² and averages 20.35Å². The average coordinate error as estimated by a plot of the crystallographic residual vs. the resolution (Luzzati, 1952) is approximately 0.21Å (see Figure 2.2.2-6). Electron density from the 2F_o-F_c map is shown in Figure 2.2.2-7. The ϕ , ψ torsion angles for each of the four subunits are shown in Ramachandran plots (Srinivasan and Parthasarathy, 1976) (Figure 2.2.2-8).

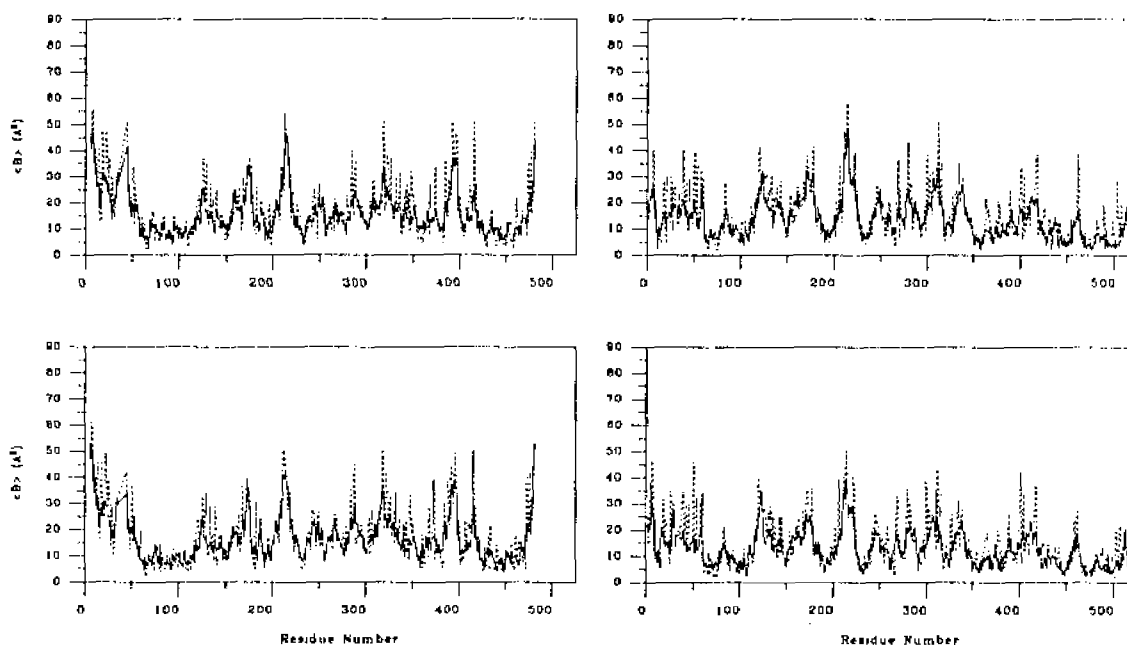


Figure 2.2.2-5. Plot of average B-factors vs. residue number for each of the four subunits of Av1 after refinement with XPLOR. The average mainchain temperature factors are shown as solid lines, and the sidechains are represented with dashed lines. Compare to Figure 2.2.2-1. The α -subunits (A and C) are shown in the upper and lower left hand panels respectively and the two β -subunits (B and D) are shown on the right.

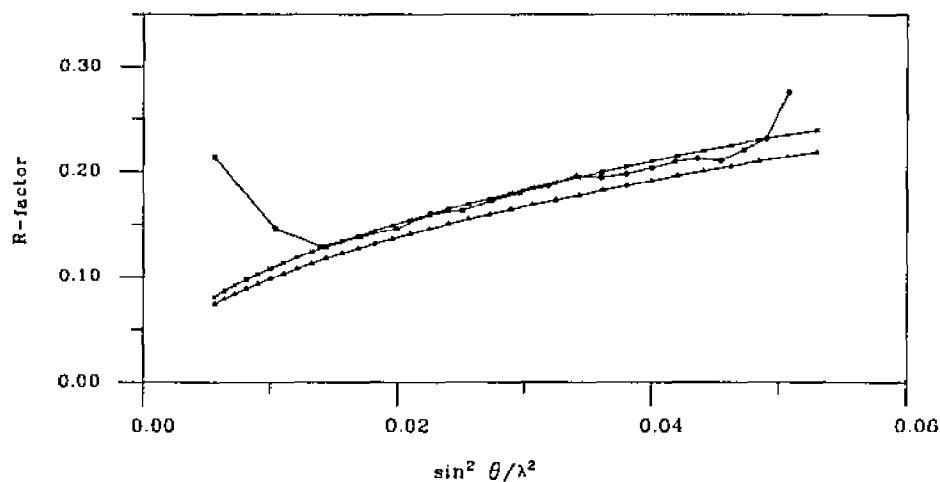


Figure 2.2.2-6. A Luzzati plot of the R-factor from the XPLOR refinement vs. the resolution. Theoretical plots are shown for an average coordinate error of 0.22 Å (top curve) and 0.20 Å (lower curve).

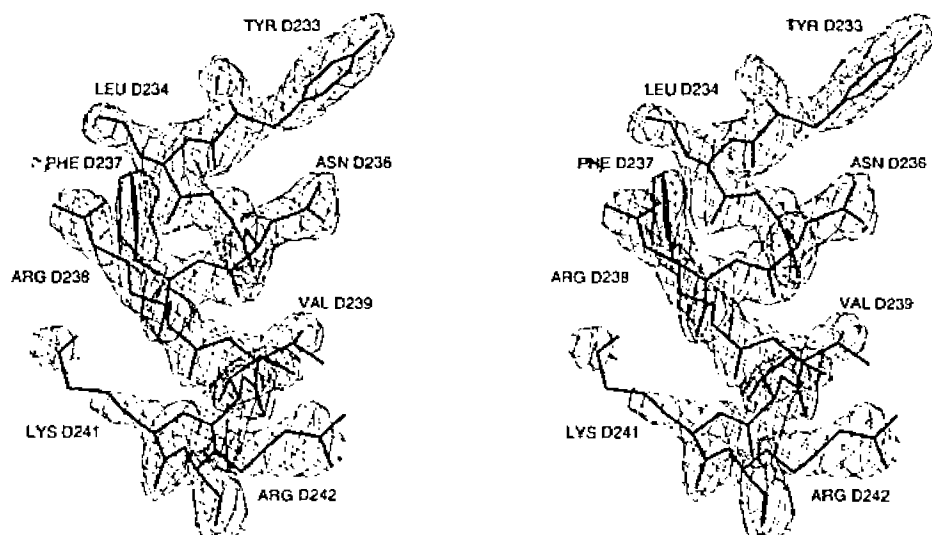


Figure 2.2.2-7. Stereo view of $2F_0-F_c$ electron density from the same helix shown in Figure 2.2.2-3. This density is representative of the entire protein after refinement with XPLOR. The map is contoured at the 2σ level.

Comparison of TNT and XPLOR Results

As observed in the refinement of rubredoxin, the range of temperature factors seen

after refinement is much larger in TNT than in XPLOR. This can be clearly seen by comparing Figures 2.2.2-1 and 2.2.2-5 where the average temperature factors for each residue (both backbone atoms and sidechain atoms have been plotted against the residue number). An interesting observation with respect to the temperature factors is that while for XPLOR refinement the average B's of the sidechains is nearly equal to that of the main chain atoms across the entire sequence, the rms. deviation in the B's of adjacent atoms is over 60% higher than for TNT refinement (2.67\AA^2 vs. 1.61\AA^2 respectively) where the side chain B's are generally much higher than the main chain average B's.

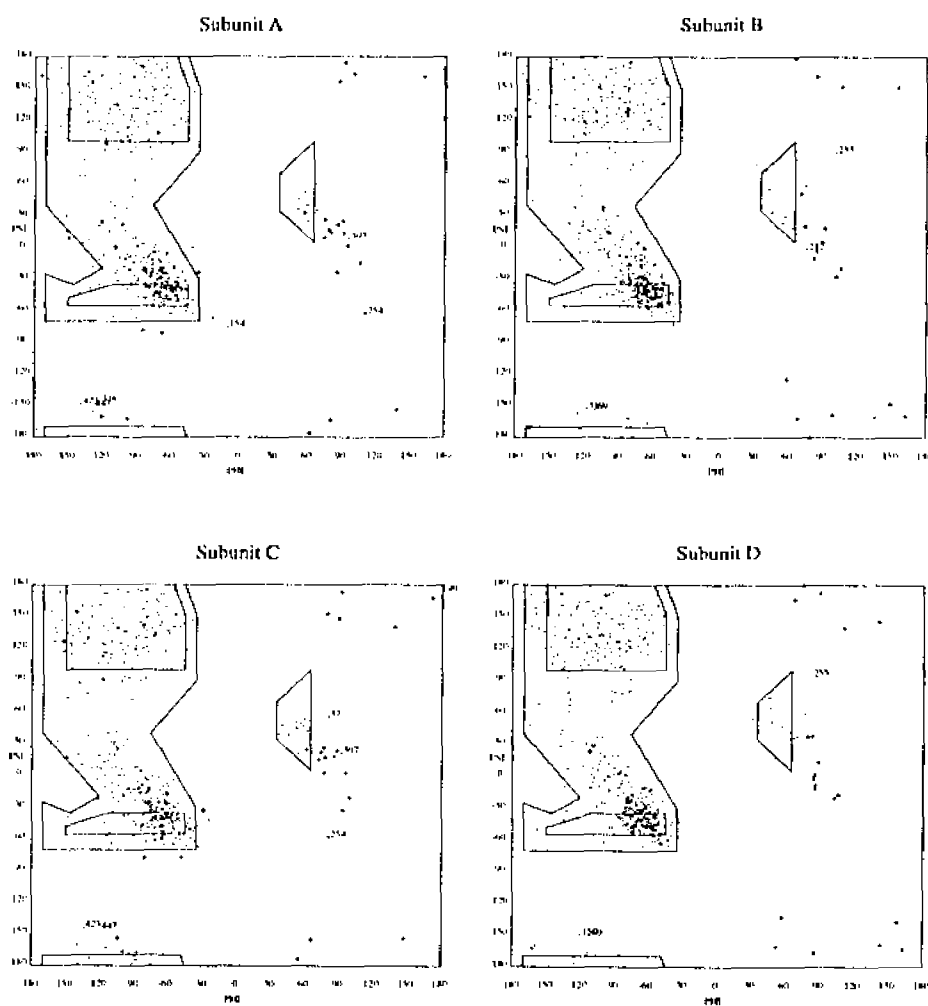


Figure 2.2.2-8. Ramachandran plot of the ϕ , ψ torsion angles for each subunit of Av1 after refinement with XPLOR. Glycine residues are represented by + and residues with ϕ , ψ angles in areas outside of the most favorable regions are labeled by residue number.

Both refinements give similar errors in bond distances and angles and the overall average coordinate error as determined by Luzzati plots are virtually the same. Inspection of Tables 2.2.2-3 thru 2.2.2-7 shows that the distances (both for bonds and interatomic distances) within the cofactors do agree well, at least within the limits of the estimated average coordinate error. However, some of the distances within the P-cluster pair, particularly those that involve S1A and S1B show large deviations.

Table 2.2.2-3. Bond distances within the M-cluster and to its ligands. The initial lengths in the right-hand column are for 1MIN. Sulfur atoms that bridge the two cluster fragments are marked with *. The atom labeling scheme is shown in Figure 2.3-9.

From	To	TNT		XPLOR		1MIN	
		α [A]	α [C]	α [A]	α [C]	α [A]	α [C]
Mo1	S1B	2.36	2.39	2.43	2.45	2.50	2.48
	S3B	2.28	2.28	2.36	2.29	2.47	2.44
	S4B	2.34	2.37	2.41	2.42	2.42	2.48
	N δ 1 α 442	2.13	2.21	2.27	2.17	2.25	2.24
	O5 CIT	2.36	2.30	2.26	2.31	2.25	2.28
	O7 CIT	2.06	2.17	2.18	2.22	2.23	2.25
Fe5	S1B	2.24	2.29	2.23	2.33	2.29	2.31
	S4B	2.27	2.22	2.38	2.26	2.30	2.32
	*S3A	2.21	2.27	2.23	2.26	2.29	2.32
Fe6	S1B	2.20	2.18	2.29	2.18	2.32	2.25
	S3B	2.25	2.24	2.19	2.24	2.31	2.26
	*S2B	2.23	2.27	2.24	2.21	2.27	2.29
Fe7	S3B	2.21	2.24	2.26	2.23	2.30	2.34
	S4B	2.26	2.21	2.42	2.33	2.37	2.34
	*S5A	2.22	2.18	2.17	2.18	1.82	1.80
Fe1	S1A	2.28	2.23	2.31	2.17	2.36	2.36
	S2A	2.28	2.20	2.31	2.22	2.38	2.30
	S4A	2.33	2.31	2.26	2.29	2.36	2.31
	S γ α 275	2.31	2.31	2.26	2.20	2.29	2.29
Fe2	S1A	2.25	2.24	2.28	2.28	2.31	2.29
	S2A	2.22	2.27	2.21	2.23	2.28	2.27
	*S2B	2.22	2.19	2.12	2.17	2.26	2.28
Fe3	S2A	2.30	2.29	2.30	2.25	2.36	2.32
	S4A	2.32	2.28	2.30	2.27	2.39	2.36
	*S5A	2.17	2.19	2.09	2.13	1.83	1.86
Fe4	S1A	2.31	2.38	2.41	2.33	2.37	2.35
	S4A	2.33	2.29	2.31	2.21	2.37	2.31
	*S3A	2.16	2.20	2.10	2.11	2.26	2.21

Table 2.2.2-4. P-Cluster bond distances. The initial bond lengths are listed in the right-hand column for the coordinates deposited in the Protein Data Bank as 1MIN. Distances for each subunit resulting from TNT and XPLOR are shown. See Figure 2.3-11 for the atom labeling scheme.

From	To	TNT		XPLOR		1MIN	
		α [A]	α [C]	α [A]	α [C]	α [A]	α [C]
Fe1	S1A	2.25	2.27	2.32	2.29	2.36	2.35
	S2A	2.24	2.33	2.11	2.33	2.28	2.37
	S4A	2.33	2.27	2.39	2.34	2.41	2.36
	S γ β 95	2.23	2.23	2.26	2.25	2.32	1.91
Fe2	S1A	2.25	2.15	2.04	2.15	2.07	2.03
	S2A	2.25	2.23	2.28	2.26	2.35	2.37
	S3A	2.23	2.31	2.26	2.37	2.29	2.36
	S γ α 154	2.35	2.30	2.32	2.28	2.62	2.11
Fe3	S2A	2.25	2.26	2.37	2.32	2.37	2.35
	S3A	2.28	2.31	2.34	2.39	2.34	2.41
	S4A	2.31	2.28	2.32	2.30	2.43	2.39
	S γ α 62	2.26	2.26	2.30	2.27	2.48	2.33
Fe4	S1A	2.24	2.27	2.22	2.29	2.31	2.29
	S3A	2.37	2.38	2.35	2.36	2.38	2.40
	S4A	2.26	2.29	2.33	2.35	2.33	2.35
	S γ α 88	2.23	2.27	2.30	2.24	2.31	2.32
Fe5	S1B	2.33	2.37	2.20	2.53	2.29	2.42
	S2B	2.25	2.32	2.24	2.28	2.30	2.41
	S4B	2.32	2.21	2.26	2.23	2.33	2.34
	S γ α 88	2.35	2.28	2.56	2.33	2.05	1.91
Fe6	S1B	2.33	2.28	2.46	2.12	2.28	2.25
	S2B	2.29	2.20	2.27	2.26	2.28	2.30
	S3B	2.35	2.31	2.36	2.43	2.39	2.39
	S γ β 153	2.25	2.29	2.22	2.28	2.58	2.56
	O γ β 188	2.30	2.35	2.17	2.22	2.16	2.17
Fe7	S2B	2.33	2.29	2.32	2.35	2.37	2.35
	S3B	2.35	2.38	2.43	2.39	2.38	2.36
	S4B	2.29	2.25	2.34	2.30	2.38	2.33
	S γ β 70	2.40	2.39	2.37	2.39	2.47	2.32
Fe8	S1B	2.12	2.18	1.73	2.03	2.10	2.20
	S3B	2.26	2.28	2.20	2.32	2.32	2.35
	S4B	2.28	2.31	2.32	2.30	2.31	2.35
	S γ β 95	2.31	2.30	2.25	2.33	2.33	2.27
S1A	S1B	1.31	1.60	1.87	2.00	2.10	2.11

Table 2.2.2-5. Selected distances within the FeMo-cofactor. Distances for both α -subunits, A and C, are listed for refinement results from TNT and XPLOR with the initial lengths from 1MIN shown for comparison.

From	To	TNT		XLPOR		1MIN	
		α [A]	α [C]	α [A]	α [C]	α [A]	α [C]
Mo	Fe5	2.80	2.79	2.76	2.69	2.91	2.91
	Fe6	2.76	2.76	2.72	2.71	2.98	2.95
	Fe7	2.58	2.65	2.70	2.64	2.84	2.92
	Fe2	5.01	5.06	5.07	5.01	5.14	5.19
	Fe3	5.16	5.12	5.12	5.07	5.33	5.28
	Fe4	5.08	5.09	5.15	5.04	5.20	5.20
	Fe1	7.09	7.12	7.09	7.08	7.38	7.34
Fe2	Fe6	2.47	2.49	2.56	2.74	2.42	2.43
Fe3	Fe7	2.72	2.62	2.57	2.55	2.61	2.57
Fe4	Fe5	2.52	2.53	2.59	2.59	2.51	2.52
Fe1	Fe5	5.02	5.07	5.03	5.13	5.24	5.22
	Fe6	5.01	4.98	5.04	5.04	5.18	5.09
	Fe7	5.04	5.06	4.98	4.99	5.14	5.17
Fe2	Fe5	3.62	3.67	3.65	3.75	3.76	3.84
	Fe7	3.49	3.61	3.57	3.54	3.50	3.75
Fe3	Fe5	3.77	3.66	3.73	3.72	4.00	3.84
	Fe6	3.77	3.71	3.74	3.77	3.96	3.83
Fe4	Fe6	3.64	3.63	3.75	3.63	3.78	3.75
	Fe7	3.63	3.60	3.67	3.54	3.74	3.77

Table 2.2.2-6. Selected distances within the P-Cluster pair. The results from TNT and XPLOR refinement for both subunits are shown in the center two columns, and the distance for PDB file 1MIN are listed on the right.

From	To	TNT		XPLOR		1MIN	
		α (A)	α (C)	α (A)	α (C)	α (A)	α (C)
Fe1	Fe5	4.22	4.16	4.45	4.10	4.41	4.25
Fe4	Fe8	4.02	3.90	4.15	3.97	4.10	3.96
Fe4	Fe5	3.19	3.07	3.22	3.13	3.14	3.07
Fe1	Fe8	2.99	3.00	3.06	2.83	3.03	2.93
Fe2	Fe6	5.71	5.91	5.75	6.10	6.18	6.27
Fe3	Fe7	7.69	7.67	7.70	7.66	7.77	7.80
Fe6	Fe5	3.19	3.19	3.50	3.52	3.42	3.38
	Fe7	2.78	2.79	2.91	2.92	2.90	2.93
	Fe8	2.95	3.03	3.16	3.41	3.25	3.30

References

Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzle, T. F. and Weng, J. (1987). Protein Data Bank. In *Crystallographic Databases - Information Content, Software Systems, Scientific Applications* (Editor), 107-132. International Union of Crystallography, Bonn/Cambridge/Chester.

- Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977). "The protein data bank: a computer-based archival file for macromolecular structures." *J. Mol. Biol.* **112**, 535-542.
- Brünger, A. T., Kuriyan, J. and Karplus, M. (1987). "Crystallographic R Factor Refinement by Molecular Dynamics." *Science* **235**, 458-460.
- Jones, A. T. (1978). "A Graphics Model Building and Refinement System for Macromolecules." *Journal of Applied Crystallography* **11**, 268-272.
- Kim, J. and Rees, D. C. (1992). "Crystallographic structure and functional implications of the nitrogenase molybdenum-iron protein from *Azotobacter vinelandii*." *Nature* **360**, 553-560.
- Luzzati, V. (1952). "Traitement Statistique des Erreurs dans la Determination des Structures Cristallines." *Acta Crystallographica* **5**, 802-810.
- Srinivasan, R. and Parthasarathy, S. (1976). Some Statistical Applications in X-ray Crystallography. In (Editor), Pergamon Press, Oxford.
- Tronrud, D. E., Ten Eyck, L. F. and Matthews, B. W. (1987). "An Efficient General-Purpose Least-Squares Refinement Program for Macromolecular Structures." *Acta Crystallographica* **A23**, 489-501.

2.2.3 Verification of M-Cluster Composition

As discussed in the introduction to this chapter (section 2.1), the structural arrangement of the metal centers of MoFe, particularly the MoFe cofactor (referred to FeMoCo or M-cluster), was an area that attracted considerable attention. The current model is shown in Figure 2.2.3-1. The initial crystallographic model for the M-cluster as proposed by Kim and Rees (Kim and Rees, 1992; Kim and Rees, 1992) contained a bridging ligand labeled as Y which was tentatively modeled as nitrogen (now labeled S 5A and modeled as sulfur). Although the other two bridging ligands (S 3A and S 2B) appeared to more consistent with sulfur, the Y-ligand showed diminished electron density and therefore its identity was more ambiguous. Additionally, a different model had been proposed by other investigators based on their initial interpretation of x-ray crystallographic data collected on the MoFe protein from *Clostridium pasteurianum* (Cp1). This alternate model contained a sulfur atom at the center of the cluster that was bonded to the six central iron atoms. This central sulfur atom was not in the Kim and Rees model, but seemed to enjoy the general support of the inorganic community in that tetrahedral geometry around iron was observed more frequently than was trigonal geometry. Spectroscopic techniques could not unambiguously determine the sulfur content and environment of FeMoCo.

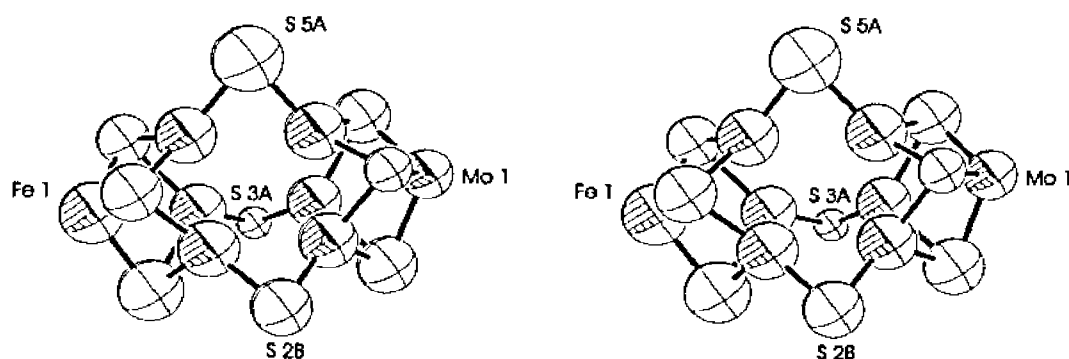


Figure 2.2.3-1. The current model of the MoFe cofactor without any of the ligands to the polypeptide chain or homocitrate. The atoms are shown as 50% probability spheres. The metal atoms are represented by the spheres with shaded octants. For clarity only the bridging sulfur atoms are labeled, and only one of the seven iron atoms are labeled.

In light of the controversy surrounding the precise composition of the MoFe-cofactor, especially with respect to the presence of a sulfur atom at the center, in addition to the identity of the bridging ligands (Orme-Johnson, 1992), an effort was made to assess the model proposed by Kim and Rees (Kim and Rees, 1992).

Y-Ligand

It was initially postulated that the Y-ligand could possibly be oxygen, nitrogen or a sulfur atom that was less well ordered than the other bridging ligands (Kim and Rees, 1992; Chan, Kim et al., 1993). An effort was made to resolve the question concerning this particular bridging ligand using TNT (Tronrud, Ten Eyck et al., 1987), the available 2.2Å resolution x-ray data and the model resulting from the first round of refinement in TNT. The methods and results are discussed below.

The temperature factors for the atoms in the M-cluster were set to the average values; i.e., 12.50Å² for the metal atoms and 11.00Å² for the sulfur atoms. The Y-ligand B-factor was set at the same value as the sulfur atoms. Two cycles of positional refinement

Table 2.2.3-1. The results of the investigation into the identity of the Y ligand.

Atom Type	Subunit	 of Sulfurs	Range of B's	σ B	B-factor of Y	Occupancy of Y
N	A	11.31	17.1 – 7.6	3.2	2.0	1.00
	C	10.88	13.8 – 4.9	2.7	2.0	1.00
O	A	11.31	17.1 – 7.6	3.2	6.51	1.00
	C	10.87	13.7 – 4.8	2.7	2.00	1.00
S	A	11.13	16.9 – 7.4	5.7	25.62	1.00
	C	10.74	13.1 – 4.6	3.9	19.84	1.00
S	A	11.68	18.3 – 6.9	3.7	14.71	0.84
	C	10.96	13.5 – 3.5	3.1	11.79	0.88

with geometric restraints were performed followed by two cycles of temperature factor refinement. Finally, the model was subjected to two cycles of combined positional and temperature factor refinement. This procedure was performed three separate times, one each with the Y-ligand as nitrogen, oxygen and sulfur with the site set at full occupancy (i.e. 1.00). The resulting temperature factors were inspected in order to evaluate the assignment of atom type for the bridging atom. The results are listed in Table 2.2.3-1.

When the bridging atom is nitrogen, the temperature factor immediately refines to the minimum value allowed by the program. When the atom type is oxygen, the B's refine to values much smaller than the average for the other atoms, with one of the oxygen temperature factors refining to the minimum. These results suggest, but don't prove, that the Y-ligand is not either oxygen or nitrogen and may indeed be sulfur. When sulfur is refined at this position, the temperature factors refine to values more in line with the other sulfur atoms but are still larger than any other sulfur.

In order to estimate the effective occupancy of the sulfur at this bridging site the above refinement procedure was repeated, but the B-factors of the Y-ligand were held constant while the occupancy of the site was to be refined during the B-refinement step. For the combined positional and temperature factor refinement step, the occupancies of all atoms were held constant and all of the B-factors were allowed to refine. These results are shown in the bottom rows of Table 2.2.3-1. With this refinement protocol, the temperature factors of the sulfur atom at this site are similar to the other sulfur atoms of the cluster, and the occupancy refines to approximately 0.85 which suggests that this site in the cluster may be more labile or mobile than the other bridging ligands. The increased lability of this site can be rationalized on the basis of its environment compared to the environment of the other bridging sulfur atoms. One of these bridging sulfurs (S 3A) sits in a pocket above four amide protons in a loop formed by Ile- α 356, Gly- α 357, Gly- α 358 and Leu- α 359 and has the lowest B-factor of all the bridging sulfurs. The other bridging sulfur (S 2B) is

surrounded by hydrophobic residues and is within H-bonding distance of His- α 195. In contrast, S 5A (the Y-ligand) is on the side of the cluster most exposed to solvent and is near Arg- α 96 and Arg- α 359. The combination of the solvent accessibility and the possible lability of this site may have implications for substrate reduction and will be discussed further in section 2.4.

For the remainder of the refinement process, the occupancy of this site was set to 1.00. The present model, averaging over both clusters and both refinement packages, seems to clearly indicate that this site is occupied by sulfur. The temperature factors for the sulfur atoms in the M-cluster range from 18.66\AA^2 to 5.02\AA^2 with an average value of 11.85\AA^2 ($\sigma=2.8\text{\AA}^2$). The temperature factor for S-5A ranges from 18.05\AA^2 to 12.29\AA^2 with an average value of 15.26\AA^2 ($\sigma=2.9\text{\AA}^2$).

It should be emphasized again that these results do not prove beyond any doubt that this site is occupied by sulfur. For example, it is not possible to exclude the presence of a phosphorous or chlorine atom at this position. However, in the absence of any compelling evidence that another atomic species other than sulfur occupies this site, it is prudent to assume that the Y-ligand is indeed a sulfur atom.

Central Sulfur

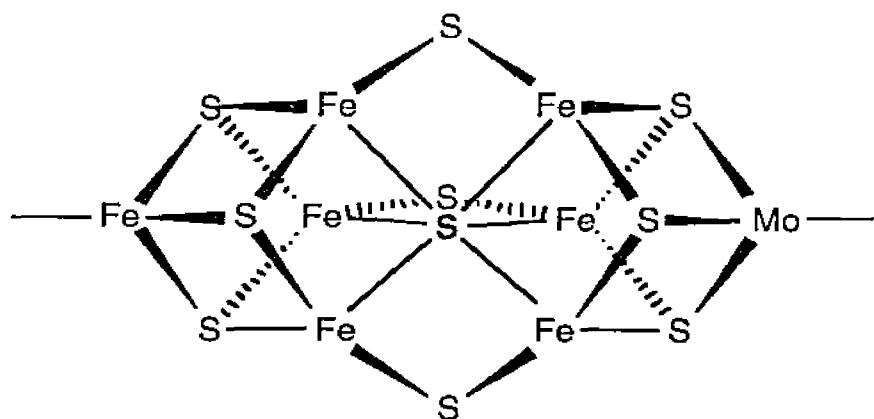


Figure 2.2.3-2. The model of the M-cluster originally proposed by Bolin et al. with a sulfur atom at the octahedral site in the middle of the cluster.

As mentioned above, another investigator had independently determined the x-ray structure of Cp1 and had proposed a similar model for the MoFe cofactor, with the significant difference that this model contained a sulfur atom in its center (see Figure 2.2.3-2). In order to test this model against our data, a sulfur atom was placed in the octahedral site at the center of the cluster. The temperature factor was set to 10.00\AA^2 , and then the position and temperature factor were refined in TNT. The restraints were adjusted so that the temperature factor of the central sulfur would be strongly correlated to the temperature factor of the atoms to which it was bonded. The geometry of the site was also restrained. After two cycles of refinement, the temperature factor of the central sulfur increased to nearly 30\AA^2 and after four cycles had settled at approximately 35\AA^2 . The large B-factor suggests that there is certainly no well ordered sulfur in this position. However, the most compelling evidence comes from the 2Fo-Fc electron density maps calculated from the refined coordinates where no density was seen for this atom. Figure 2.2.3-3 shows the electron density from the 2Fo-Fc map of the M-cluster in the most current model.

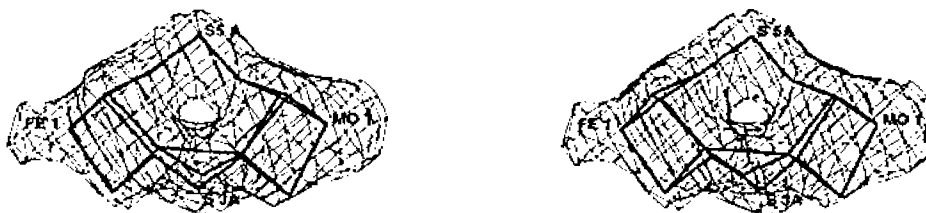


Figure 2.2.3-3. Stereo view of the electron density from the 2Fo-Fc map from the current model of the M-cluster after refinement with TNT. The map is contoured at the 3σ level. A hole is clearly visible in the center of the cluster. Bridging ligand S-2A is at the rear of the figure and is not labeled.

This result serves to emphasize the unique nature of this cluster. Taken together with the position of the molybdenum atom, it reiterates the potential pitfalls of incorporating preconceived beliefs into the analysis of x-ray crystallographic data, especially when working at less than atomic resolution.

References

- Chan, M. K., Kim, J. and Rees, D. C. (1993). "The Nitrogenase FeMo-Cofactor and P-Cluster Pair: 2.2Å Resolution Structures." *Science* **260**, 792-794.
- Kim, J. and Rees, D. C. (1992). "Crystallographic structure and functional implications of the nitrogenase molybdenum-iron protein from *Azotobacter vinelandii*." *Nature* **360**, 553-560.
- Kim, J. and Rees, D. C. (1992). "Structural Models for the Metal Center in the Nitrogenase Molybdenum-Iron Protein." *Science* **257**, 1677-1682.
- Orme-Johnson, W. H. (1992). "Nitrogenase Structure - Where to now?" *Science* **257**, 1639-1640.
- Tronrud, D. E., Ten Eyck, L. F. and Matthews, B. W. (1987). "An Efficient General-Purpose Least-Squares Refinement Program for Macromolecular Structures." *Acta Crystallographica* **A23**, 489-501.

2.2.4 Verification of P-Cluster Composition

The exact structure of the P-cluster pair is still an open question. Our current model for the P-cluster pair from Av1 is shown in Figure 2.2.4-1. The model is essentially the same model proposed by Chan and Rees (Chan, Kim et al., 1993). This model differs only slightly from the original Kim and Rees model (Kim and Rees, 1992) by the presence of a disulfide bond joining two of the central sulfur atoms, which are designated S 1A and S 1B, in the current model. Interestingly, these sulfurs are the very site of disagreement between the model proposed by Bolin et al. for Cp1 and the model shown below. The latest model proposed by Bolin (Bolin, Campobasso et al., 1993) contains a single sulfur atom bridging the two iron atoms, Fe(2) and Fe(6). A number of refinement experiments were conducted in order to distinguish between the two possibilities. The model proposed by Bolin et al. not only has one sulfur atom at the di-sulfide linkage site, but would also require significantly different Fe atomic positions for Fe(2) and Fe(6), as a consequence of having only one sulfur atom bridging between them. That is to say if only one sulfur atom

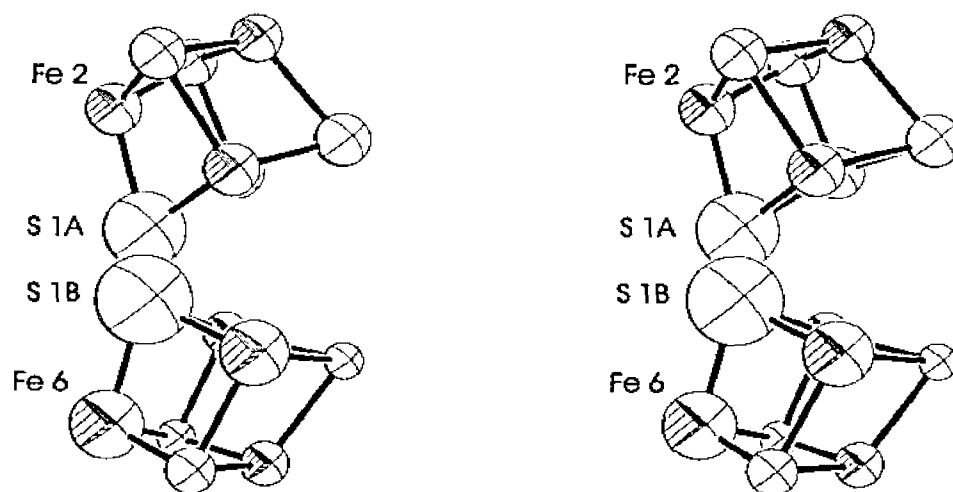


Figure 2.2.4-1. Stereo view of the current model for the P-cluster pair. The atomic spheres represent 25% probability electron density. The metal atoms are shown as spheres with shaded octants, and the sulfur atoms are represented by spheres with unshaded octants.

bridges Fe(2) and Fe(6), then the distance between these two iron sites would be approximately 5Å or less and not the nearly 6.2Å observed for the model listed in the PDB.

The structure of the P-cluster pair was assessed by removing the P-cluster pair and all of its ligands from the model and then calculating a delete map, both before and after refinement. This approach was used for both the penultimate model refined with TNT (including solvent oxygens) and for the initial model built from the experimental map and never refined against data.

Refined Coordinates

The P-cluster pair and its ligands were removed from the penultimate set of coordinates, which included 626 solvent oxygen atoms. Difference maps (2Fo-Fc and Fo-Fc) were calculated from this model (without refinement) using all data between 25-2.2Å. The Fo-Fc map was searched for the highest peaks using the peak searching

Table 2.2.4-1. The atomic positions for the Fe atoms of the P-cluster pairs in the penultimate model are listed in the first column for comparison. The positions of the highest peaks in the Fo-Fc map of the P-cluster deleted model are in the second column and the highest peaks in the Fo-Fc maps after refinement of the deleted model are listed in the right column.

Atom	Chain	Fe Coordinates			Difference Peaks			Difference Peaks		
		Complete Model			Deleted Model			Refined Delete		
		X	Y	Z	X	Y	Z	X	Y	Z
Fe1	A	-7.897	-7.814	2.479	-8.00	-7.88	2.69	-7.99	-7.86	2.64
Fe2	A	-9.730	-6.669	1.054	-9.82	-6.55	1.10	-9.72	-6.50	1.07
Fe3	A	-8.741	-9.016	0.112	-8.65	-9.24	0.01	-8.63	-9.24	0.03
Fe4	A	-7.249	-6.800	0.012	-7.07	-6.67	-0.02	-7.10	-6.70	0.00
Fe5	A	-4.864	-5.022	1.172	-4.89	-4.94	0.86	-4.88	-4.92	0.89
Fe6	A	-6.020	-2.825	3.214	-6.35	-2.60	3.15	-6.31	-2.62	3.18
Fe7	A	-3.717	-4.366	3.590	-3.57	-4.29	3.63	-3.58	-4.28	3.60
Fe8	A	-5.984	-5.761	3.730	-6.02	-5.86	3.78	-6.00	-5.85	3.78
Fe1	C	62.081	-26.504	2.499	62.10	-26.46	2.72	62.12	-26.46	2.67
Fe2	C	63.995	-27.588	1.179	64.24	-27.79	1.34	64.13	-27.80	1.20
Fe3	C	62.996	-25.244	0.262	62.93	-25.04	0.12	62.88	-25.06	0.13
Fe4	C	61.472	-27.494	0.132	61.26	-27.68	-0.13	61.27	-27.69	-0.13
Fe5	C	59.103	-29.244	1.323	59.21	-29.20	1.14	59.26	-29.16	1.14
Fe6	C	60.126	-31.536	3.272	60.43	-31.72	3.28	60.42	-31.69	3.29
Fe7	C	57.895	-29.903	3.638	57.58	-29.96	3.62	57.65	-30.01	3.56
Fe8	C	60.233	-28.557	3.703	60.15	-28.55	3.70	60.38	-28.55	3.75

program PEAKMAXR. The maps were also visually inspected. The eight highest peaks in the Fo-Fc map corresponded to the eight Fe atomic positions in the P-cluster model within the estimated coordinate error (see Table 2.2.4-1 and Figure 2.2.4-2). Visual inspection of the 2Fo-Fc map contoured at 6σ (not shown) indicated strong density for seven of the sulfur positions and weaker density for sulfur S 1B. This map contoured at 8σ shows density for the disulfide bridge (see Figure 2.2.4-3) although no density appears at the position occupied by S 1B. These results, i.e., the Fe positions and sulfur density are consistent with the model.



Figure 2.2.4-2. Stereo view of the P-cluster pair showing all peaks in the Fo-Fc map above the 12σ level and superimposed on the coordinates of the deleted atoms.

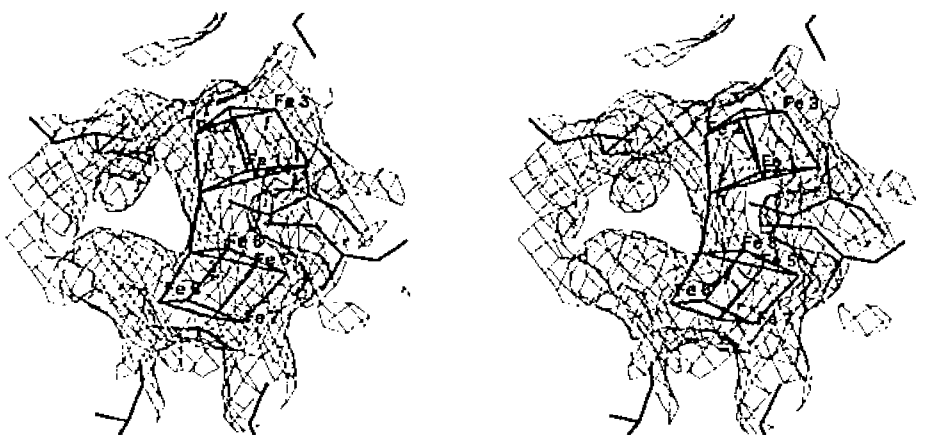


Figure 2.2.4-3. Stereo view of the 8σ electron density of the 2Fo-Fc map calculated from the same model used in the map shown in Figure 2.2.4-1 above. The P-cluster pair and ligand coordinates are superimposed on map for reference.

This P-cluster deleted model was then refined with the program TNT against the

data set collected at SSRL, using all of the reflections between 25-2.2Å resolution.

Refined delete difference maps (2Fo-Fc and Fo-Fc) were calculated as above and the Fo-Fc map was searched with PEAKMAXR as before. The eight highest peaks in the Fo-Fc map were at positions virtually identical to those in the unrefined delete maps (see Table 2.2.4-1 and Figure 2.2.4-4). The 2Fo-Fc map was visually inspected and, as before, the 2Fo-Fc map at the 6σ level indicated strong density for seven of the eight sulfur atoms and weaker density for sulfur S 1B. The 2Fo-Fc map contoured at 8σ is shown in Figure 2.2.4-5 and again shows density for the disulfide bridge but no density for S 1B.

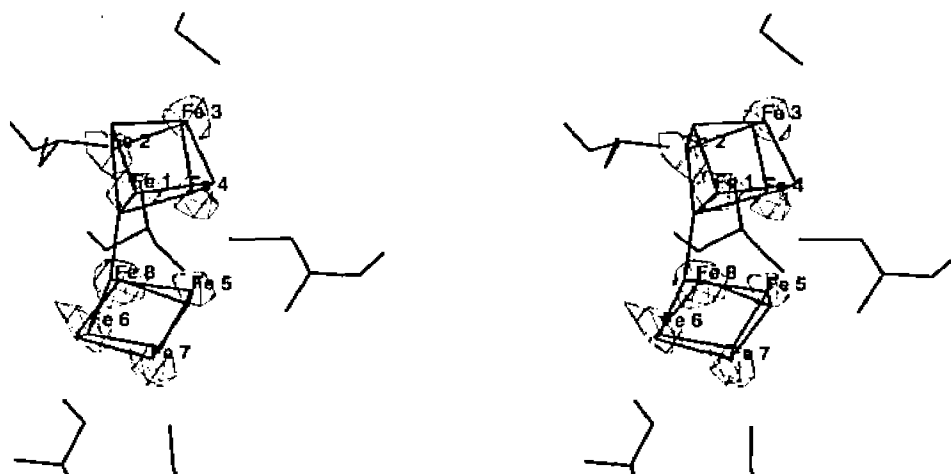


Figure 2.2.4-4. Stereo view of the peaks above 12σ in the Fo-Fc map calculated using the model with the P-cluster or its ligands after refinement with TNT. The map is superimposed on the coordinates of the deleted atoms.

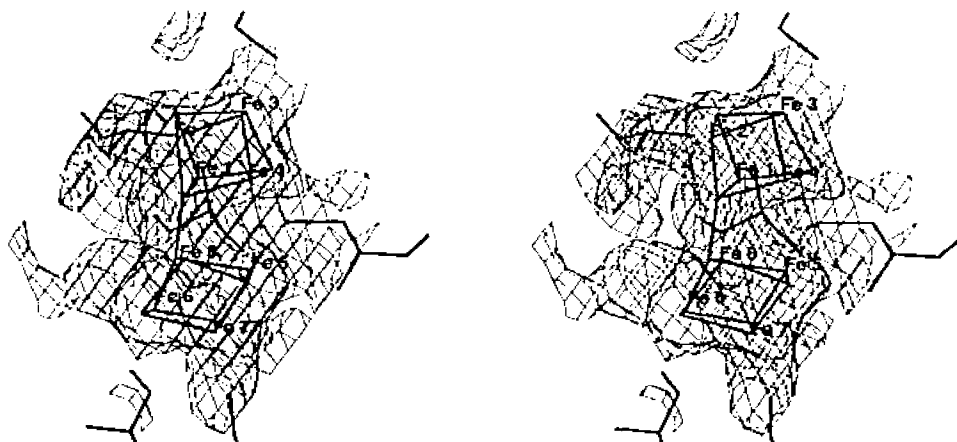


Figure 2.2.4-5. Stereo view of the 2Fo-Fc electron density map from the same coordinates as used above. Contours correspond to 8σ electron density.

Unrefined Coordinates

Although the experiments discussed and the maps shown above evidently confirm the model as proposed, it was necessary to assure that the refined model was not introducing bias into the results. Therefore, the P-cluster pair was completely removed (as before) from the coordinates that were built directly into the experimental map. These coordinates had not been refined against any data set. The model contains no solvent oxygen atoms. The data set used at this stage was a merged data set that included data collected in house (maximum resolution 2.7Å) and data collected at SSRL (2.2Å maximum resolution). Delete difference maps (2Fo-Fc and Fo-Fc) were calculated. All the reflections between 25-2.2Å resolution were used in the calculation. As before the Fo-Fc map was searched with the program PEAKMAXR. The eight Fe positions were not the strongest features of the map, but a visual inspection of the map verified the eight Fe atomic positions with electron density peaks at 12 times the standard deviation of the map (not

Table 2.2.4-2. The original iron positions are listed in the left-hand column and the positions of the peaks from the difference maps after refinement of the model against the combined data set are shown in the right-hand column.

		Original Iron Positions			Refined Peak Positions Deleted Model		
Atom	Chain	X	Y	Z	X	Y	Z
FE1	A	-7.939	-8.036	2.357	-7.81	-7.87	2.52
FE2	A	-9.973	-7.061	0.791	-9.81	-6.46	0.94
FE3	A	-8.598	-9.349	0.019	-8.79	-9.18	0.06
FE4	A	-7.301	-6.837	-0.072	-7.10	-6.88	-0.09
FE5	A	-5.076	-4.426	1.300	-4.80	-5.25	1.47
FE6	A	-5.394	-2.806	3.505	-6.28	-2.61	3.29
FE7	A	-3.188	-4.477	3.307	-3.64	-4.22	3.62
FE8	A	-5.757	-5.584	3.738	-6.10	-5.99	3.86
FE1	C	62.180	-26.271	2.444	62.04	-26.44	2.68
FE2	C	64.229	-27.246	0.888	64.05	-27.77	1.20
FE3	C	62.860	-24.959	0.114	62.95	-25.01	0.06
FE4	C	61.565	-27.471	0.021	61.19	-27.77	-0.10
FE5	C	59.328	-29.882	1.385	59.31	-29.17	1.15
FE6	C	59.626	-31.501	3.586	60.53	-31.66	3.26
FE7	C	57.421	-29.830	3.383	57.79	-29.93	3.51
FE8	C	59.986	-28.723	3.818	60.36	-28.56	3.80

shown).

The coordinates were then refined with TNT against the combined data set using all of the reflections between 25-2.2Å. A difference map was calculated at the resolution limit of the refinement and searched for strong peaks as before. The positions of the strongest peaks coincide with the positions of the iron atoms in the P-cluster pair (from the unrefined coordinate set) and are listed in Table 2.2.4-2.

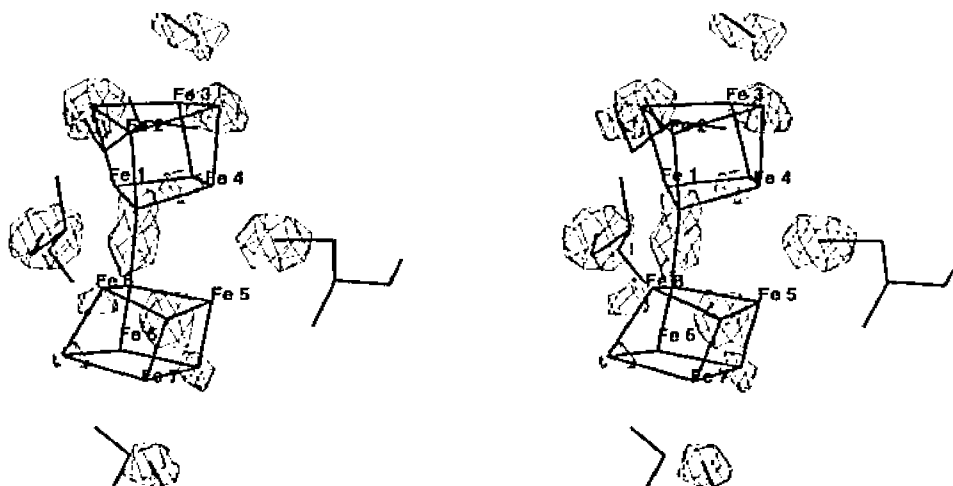


Figure 2.2.4-6. Stereo view of the Fo-Fc electron density map after refinement of the model with the iron atoms of the P-cluster pair. The density is shown at the 6σ level.

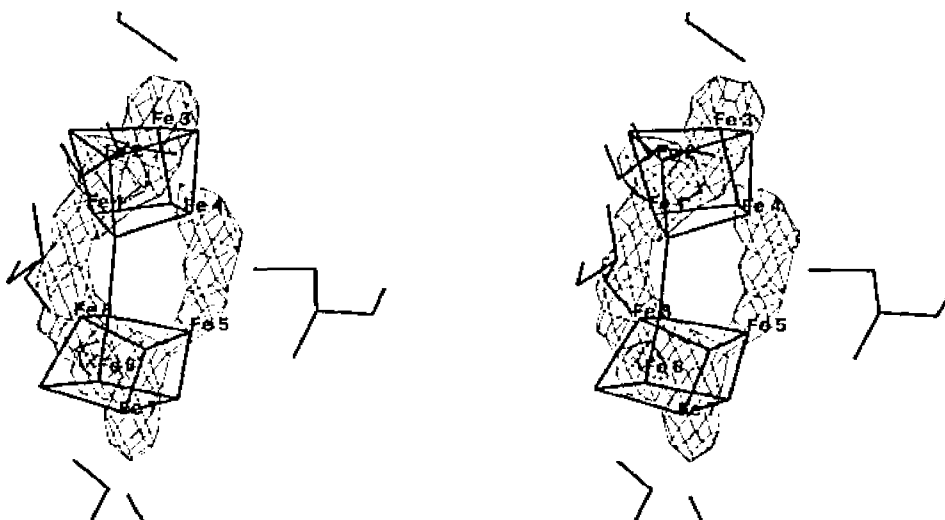


Figure 2.2.4-7. Stereo view of the electron density at 5σ of the $2F_o-F_c$ map from the refinement above. The map is superimposed on the P-cluster and its ligands.

The Fe atoms of the P-cluster were placed at the positions indicated from the eight

highest peaks in the difference map (calculated after refinement of the initial coordinates against the merged data with the entire P-cluster pair deleted). The resulting coordinates were then refined against the merged data. Maps were calculated at both the resolution limit of the refinement and at a lower resolution of 2.8\AA . Figure 2.2.4-6 shows the Fo-Fc map contoured at the 6σ level and clear density is evident for all of the sulfur positions and for the disulfide bond. Consistent with the previous results, the density for S 1B is weak. Figure 2.2.4-7 shows the 2Fo-Fc map at the 5σ level. The iron positions are consistent with the electron density. An Fo-Fc map was calculated to a maximum resolution of 2.8\AA using the same set of coordinates used for the maps shown in Figures 2.2.4-6 and 2.2.4-7. This map is shown in Figure 2.2.4-8. Electron density is evident for all of the sulfur atoms. However, a striking feature of this map is the strong spherical density on the sulfur to sulfur vector between Fe(2) and Fe(6) (compare to Figure 2.2.4-6). It is notable that this density is displaced toward the position that would be occupied by S 1A.

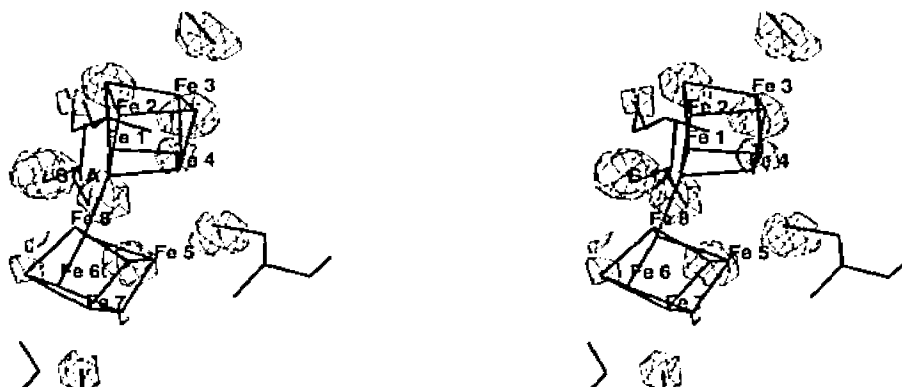


Figure 2.2.4-8. Stereo view of the Fo-Fc map at 6σ and 2.8\AA . Significant density can be seen between the positions that would be occupied by S 1A and S 1B.

The above experiment was repeated using the 2.7\AA resolution data collected in house on the area detector. The electron density from the Fo-Fc map after refinement of the model with the iron atoms included is shown in Figure 2.2.4-9. Once again a strong difference peak is observed along the bond vector between S 1A and S 1B. It is significant to note that no density appears for this position in the 2Fo-Fc map at 2σ even

though all of the other sulfur positions in the P-cluster pair do appear in this map (see Figure 2.2.4-10).

The experiments done to this point seem to firmly establish the atomic positions of the iron atoms for the P-cluster pair. As discussed above, a model containing a single hexacoordinate sulfur atom as a bridge would require significantly different distances between Fe(2) and Fe(6) than those observed here. The average distance between Fe(2) and Fe(6) as calculated from the positions of the refined difference peaks listed in Tables 2.2.4-1 and 2.2.4-2 is 5.7\AA . This would require a sulfur to iron bond distance in excess of 2.8\AA . This requirement exceeds the sum of the generally accepted covalent radii for iron

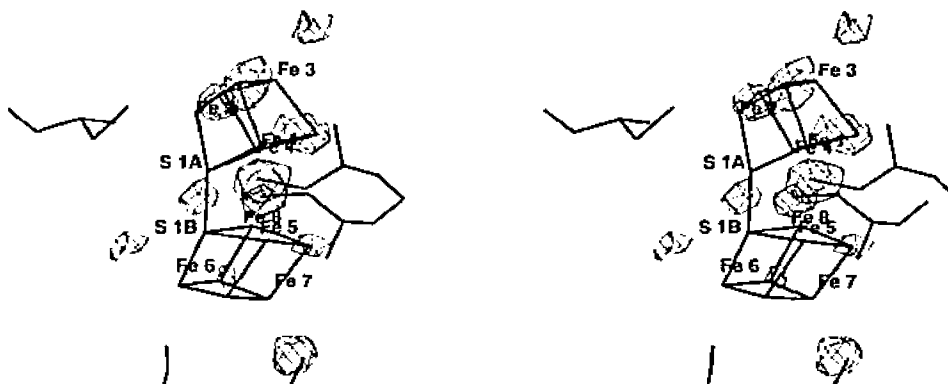


Figure 2.2.4-9. Stereo view of the F_o-F_c electron density at 6σ after refinement with the 2.7\AA data collected in house. Significant electron density can be seen along the S(1A) and S(1B) bond vector.

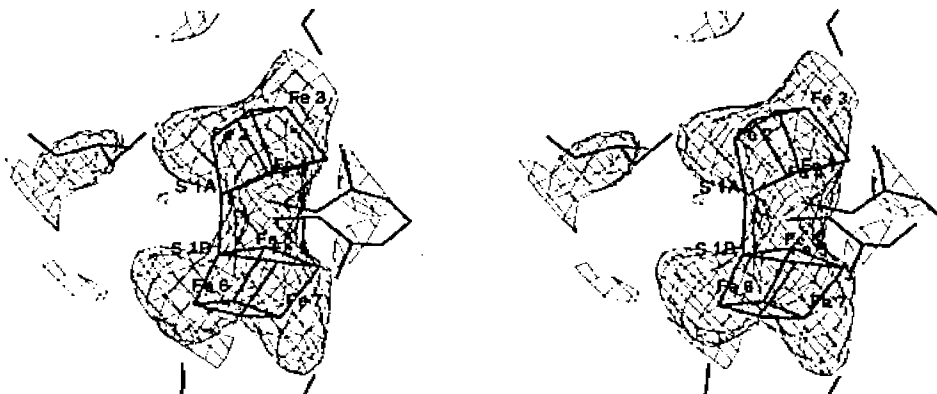


Figure 2.2.4-10. Stereo view of the electron density from the $2F_o-F_c$ map at 2σ . This map was calculated after refinement against the low resolution data set. Notice the absence of electron density between S(1A) and S(1B).

(1.17Å) and sulfur (1.02Å). To our knowledge an iron-sulfur bond of this length has not been observed. The current model has an Fe(2) to Fe(6) distance of approximately 5.85Å.

Additionally, it would appear from the above results that the positions of the sulfur atoms for the P-cluster ligands are correctly assigned. The sulfur atoms of the P-cluster pair themselves are well behaved with the exception of S 1A and S 1B.

Unrefined Coordinates (Single sulfur deleted)

In light of the somewhat ambiguous results from the previous refinement experiments with respect to the presence or absence of S(1A) or S(1B), additional

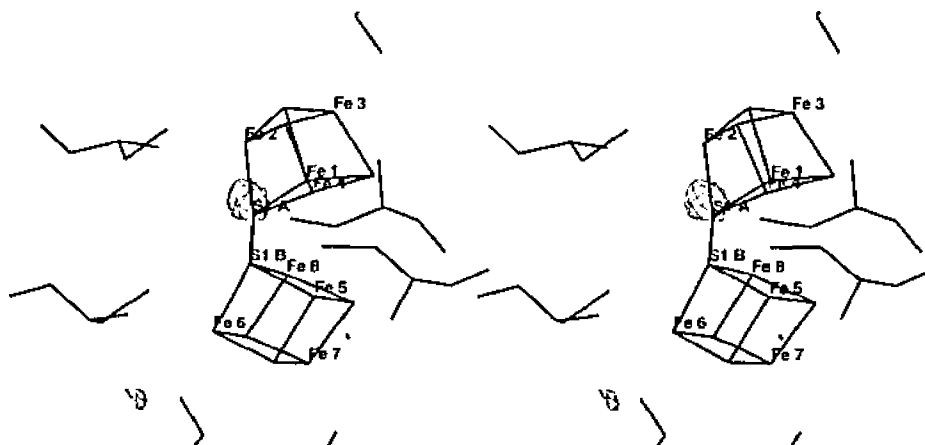


Figure 2.2.4-11. Stereo view of 5 σ electron density from the Fo-Fc map after S 1A is deleted and the coordinates are refined.

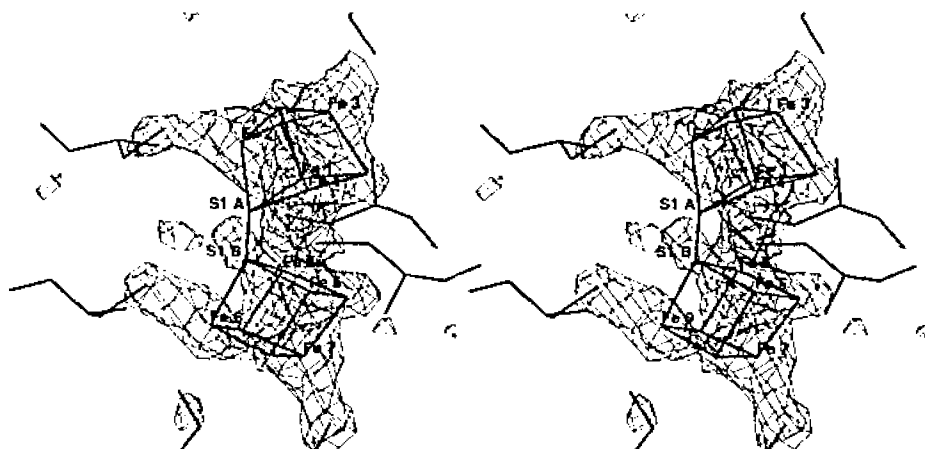


Figure 2.2.4-12. Stereo view of electron density from the 2Fo-Fc map around the P-cluster pair after deletion of S 1A and refinement of the model. The map is contoured at the 4 σ level.

experiments were done focusing on these two atoms. Each of these two atoms were removed from the unrefined model separately and the model was then refined against the combined data set using TNT and all reflections between 25.0 - 2.2Å resolution. Figure 2.2.4-11 shows the difference electron density at 5 σ after deletion of S 1A and refinement with TNT. The peak for S 1A is clearly the dominant feature of the map. The electron density from the 2Fo-Fc map is shown in Figure 2.2.4-12. This map shows 4 σ density for all of the atoms in the P-cluster pair except S 1A. Strong density is evident along the S 1A – S 1B bond vector. A striking feature of this map is the lack of electron density in the center of the 4Fe4S cubane containing Fe(5), Fe(6), Fe(7) and Fe(8). A similar trend is seen in all of the high resolution 2Fo-Fc maps shown this far.

The results obtained when S 1B is deleted from the model prior to the refinement are shown in Figures 2.2.4-13 and 2.2.4-14. The difference map (Figure 2.2.4-13) shows 5 σ electron density at the position corresponding to S 1B. As before the 2Fo-Fc map (Figure 2.2.4-14) shows strong density (>4 σ) for all the atomic positions of the P-cluster pair except the atom that was deleted. The "hole" in the center of the cubane is observed once again. This is a puzzling and as yet unexplained feature of the maps.

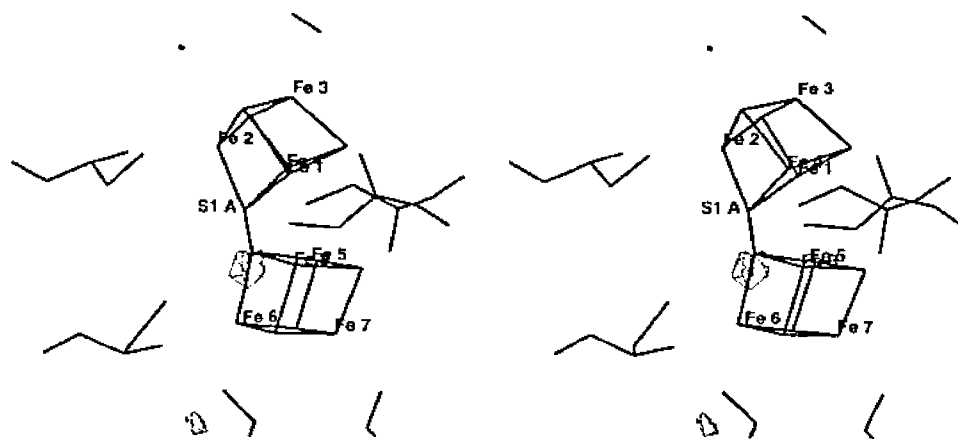


Figure 2.2.4-13. Stereo view of 5 σ electron density from the Fo-Fc map after S(1B) is deleted and the coordinates are refined.



Figure 2.2.4-14. Stereo view of electron density from the $2F_o - F_c$ map around the P-cluster pair after deletion of S(1A) and refinement of the model. The map is contoured at the 4σ level.

Electron Density Averaging

The MoFe tetramer is a dimer of dimers related by non-crystallographic symmetry (an approximate two-fold rotation parallel to the Z-axis). It is possible to exploit this symmetry and use non-crystallographic symmetry averaging of the electron density to further investigate the composition of the P-cluster pair. The program RAVE (Kleywegt and Jones, 1994) was used for the averaging. All map calculations used the CCP4 (SERC, 1986) suite of programs. The mask used for the averaging was calculated with the program MAMA (Kleywegt and Jones, 1994). All map format conversions and manipulations were done with the program MAPMAN (Kleywegt and Jones, 1994).

Forty cycles of averaging were performed and convergence was obtained within these 40 cycles. Each cycle was composed of the following steps: (1) inversion of the input map to obtain calculated structure factors, (2) scaling of F_{obs} and F_{calc} then calculation of an R-factor between the observed and newly calculated structure factors, (3) calculation of a $2F_{obs} - F_{calc}$ map for use in the averaging cycle, (4) extension of the new map to include slightly more than the asymmetric unit in order to facilitate reliable interpolation of electron density values at grid points near the edge of the asymmetric unit, and (5) averaging of the electron density inside the protein mask and the electron density within the area defined by

the rotation and translation of that mask consistent with the non-crystallographic symmetry of the asymmetric unit. All map calculations (including the inversion step) were done using the proper space group symmetry for the unit cell. The final map produced by the averaging cycles was a $2F_o - F_c$ map that could then be viewed (after being converted to the format required by TOM).

The penultimate model refined with TNT was used for electron density averaging. The model was modified by deleting the entire P-cluster pair, and all ligands to the P-cluster were truncated to alanine. Maps calculated with F_{obs} and α_{calc} , where α_{calc} is the calculated phase angle for an observation (F_{obs}), were used as the initial input for averaging. The resulting $2F_o - F_c$ map is shown in Figure 2.2.4-15 with electron density contours corresponding to 5σ . Strong density is observed for all of the atomic positions corresponding to iron as well as a finger of density pointing toward the atomic position for S(1A).



Figure 2.2.4-15. Stereo view of $2F_o - F_c$ electron density map after averaging of electron density with the entire P-cluster pair and all liganding atoms deleted from the model. The map is contoured at 5σ .

For the next averaging experiment only the sulfur atoms of the P-cluster pair were removed. The resulting averaged electron density map is shown in Figure 2.2.4-16 contoured at the 3σ level. Strong density is apparent for all of the deleted sulfur atoms except S 1B. Nevertheless, these results indicate that the model is essentially correct although the position for S 1B is tenuous.



Figure 2.2.4-16. Stereo view of 2Fo-Fc electron density map after averaging of electron density with the sulfur atoms of the P-cluster pair deleted from the model. The map is contoured at 3σ .

References

- Bolin, J. T., Campobasso, N., Muchmore, S. W., Morgan, T. V. and Mortenson, L. E. (1993). Structure and Environment of Metal Clusters in the Nitrogenase Molybdenum-Iron Protein from *Clostridium pasteurianum*. Molybdenum Enzymes, Cofactors, and Model Systems. ACS Symposium Series 535, American Chemical Society, Washington D.C.; pp. 186-195.
- Chan, M. K., Kim, J. and Rees, D. C. (1993). "The Nitrogenase FeMo-Cofactor and P-Cluster Pair: 2.2Å Resolution Structures." *Science* **260**, 792-794.
- Kim, J. and Rees, D. C. (1992). "Structural Models for the Metal Center in the Nitrogenase Molybdenum-Iron Protein." *Science* **257**, 1677-1682.
- Kleywegt, G. J. and Jones, T. A. (1994). "Convenient Single and Multiple- Crystal Real-Space Averaging of Macromolecular Electron- Density Maps." *To be published*.
- SERC, D. L. (1986). "CCP4. A Suite of Programs for Protein Crystallography." *SERC Daresbury Laboratory*, Warrington, England,

2.3 Structure Discussion

The nomenclature convention used in the following discussion of the structure will be the same as that adopted earlier by Kim and Rees (Kim and Rees, 1992) with one addition. Secondary structural elements of each subunit will be designated by H for the alpha helices and S for the strands of the β -sheets. The letter will be immediately followed by a number designating the position of that element in the primary sequence relative to the other elements of the same type. This will be followed by the subunit designation in square brackets. For example H3[β] designates the third alpha-helix of the beta-subunit and S7[α] represents the seventh beta-strand of the alpha-subunit. The β -sheets will be named according to the domain they are in.

Protein Fold

The MoFe-protein component of the nitrogenase system exists as an α_2/β_2 tetramer with a total molecular weight of approximately 240 kDa. The α and β subunits have nearly identical folds and are approximately the same size, with 491 and 522 amino acids respectively. The respective sequences are encoded by the *nifD* and *nifK* genes. Each subunit is composed of three domains (noted as I, II and III for the α -subunit and I', II' and III' for the β -subunit) and each domain is of the α -helical/ β -sheet type. This general structural motif is quite common in proteins, but the particular fold observed for the MoFe proteins has not been previously observed to our knowledge.

Each of the three domains of the α -subunit consist of a central β -sheet flanked by α -helices as shown in Figure 2.3-1. A schematic diagram of the α -subunit is shown in Figure 2.3-2. Domain I consists of seven helices (designated by H3[α]-H9[α] in the sequence) and five stranded β -sheet. The β -sheet is composed of four parallel strands (S5[α], S4[α], S2[α] and S3[α] in order) with one anti-parallel β -strand hydrogen bonded to S3[α]. The antiparallel strand (S1[β]) is contributed to the sheet by the β -subunit. This

gives an overall order of strands in sheet I of S4[α], S3[α], S1[α], S2[α] and S1[β]. The β -sheet of domain I is sandwiched between the five main α -helices in the same domain. Helices H6[α] and H8[α] are located on the more solvent exposed side of the sheet and α -helices H3[α], H9[α] and H4[α] are oriented on the more interior side of the sheet. Domain II contains a four stranded parallel β -sheet (strands S7[α], S6[α] S8[α] and S9[α] in order) surrounded by four helices (H10[α] thru H13[α]). The two α -helices H10[α] and H13[α] are on the exterior side of the sheet with H11[α] and H12[α] toward the core of the subunit. Domain III contains a six stranded β -sheet (S14[α], S13[α], S10[α], S11[α], S12[α] and S1[α]) and eight helices (H1[α], H2[α] and H15[α]-H20[α]). Sheet III is primarily a parallel β -sheet, but one strand (S1[α]) on the edge runs antiparallel to the others. The helices H17[α] and H18[α] are located on the exposed side of sheet III while helices H15[α], H16[α] and H19[α] help form the domain core. Domain III contains a short left-handed helix (H19[α]) near the M-cluster binding site. The longest helix in the α -subunit (H14[α], containing residues α 318- α 345) spans both domains II and III.

The domain structure of the β -subunit is essentially the same as that of the α -subunit (see Figures 2.3-3 and 2.3-4). There are no antiparallel strands in the β -sheets and the N-terminal end of the polypeptide chain contains four helices before the first strand, instead of only two. The N-terminus of the α -subunit is associated with domain III while that association is not established in the β -subunit. In place of this interaction, S1[β] is hydrogen bonded to the parallel β -sheet in domain I of the α -subunit. Another feature of the β -subunit is the presence of an additional short α -helix before S7[β] in domain II. Additionally, the α -L helix of the α -subunit (H19[α]) is a right-handed helix (H22[β]) in the β -subunit.

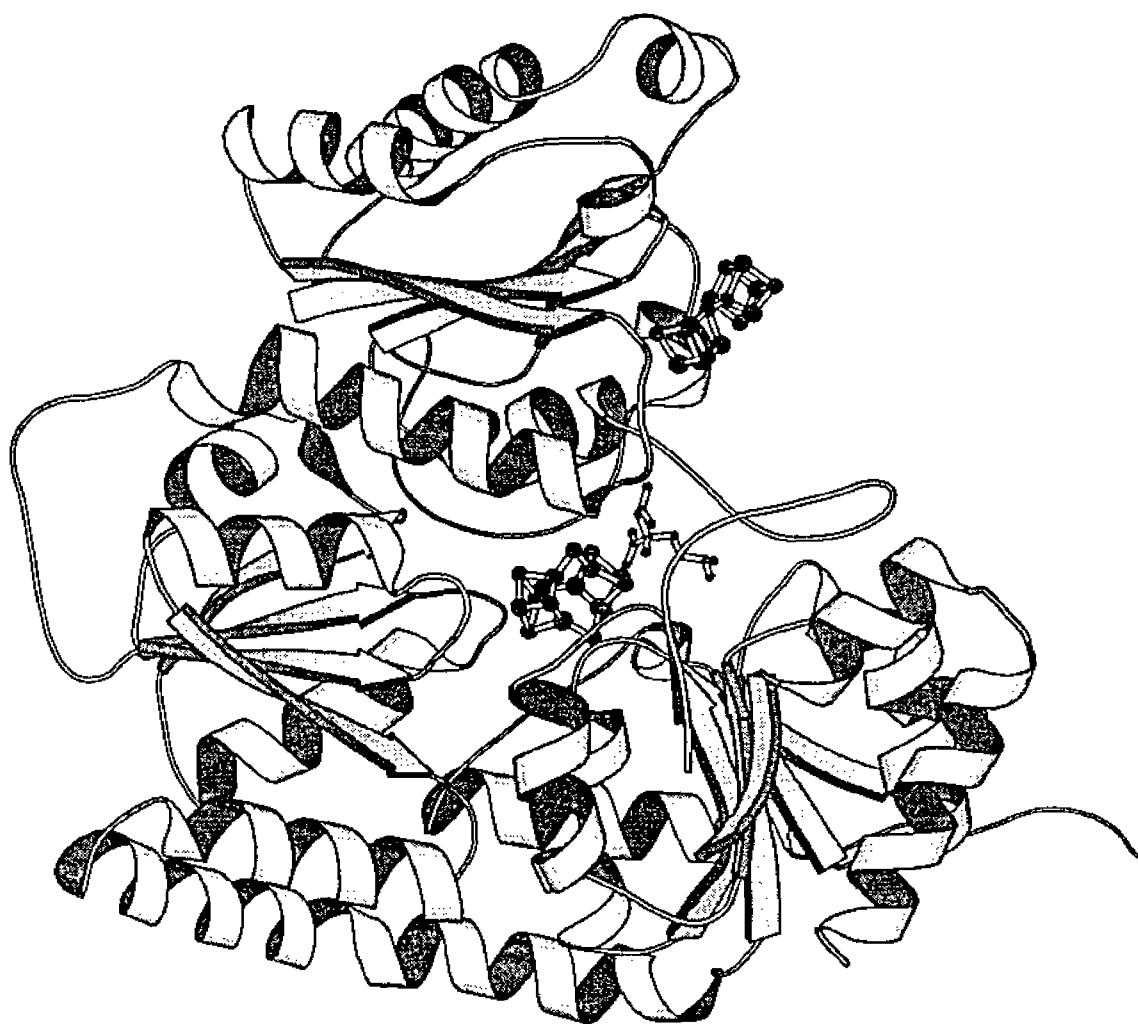


Figure 2.3-1. Molscript drawing of the α -subunit of Av1 showing the three domains, I, II and III. Domain I and the P-cluster pair at the top of the molecule. The MoFe-cofactor and the homocitrate are in the middle between domain II on the left and domain III (with its one anti-parallel β -strand) on the right. Domains II and III contribute ligands to the FeMo-cofactor (see Figure 2.3-2). The helix that spans domains II and III (H14[α]) is at the extreme bottom of the figure.

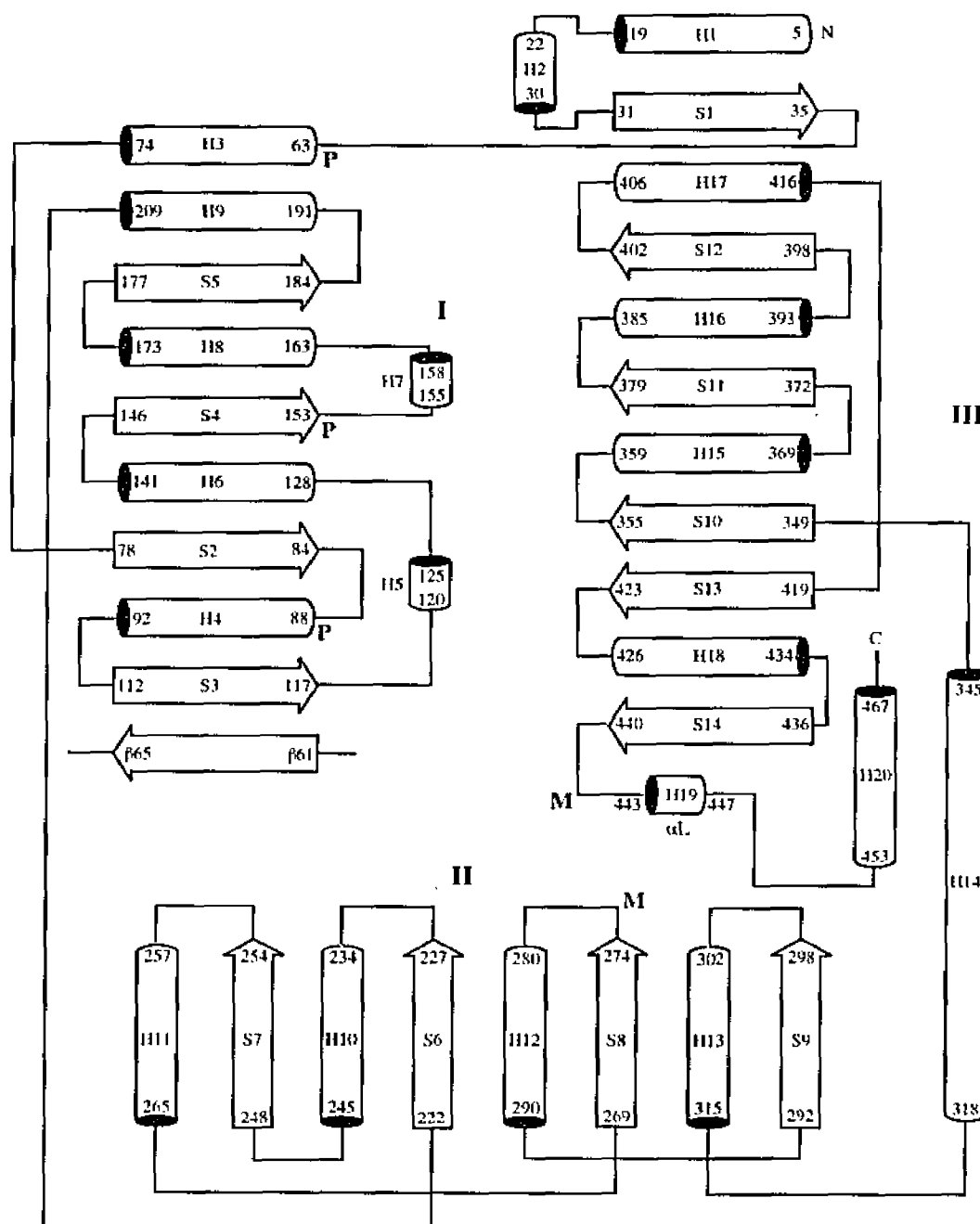


Figure 2.3-2. A schematic diagram of the secondary structure of the MoFe α -subunit. The three domains are labeled I, II and III. The positions of ligands to the M-cluster and the P-cluster are designated by M and P respectively. Within each domain, the shaded helices are on the same side of the β -sheet and are situated towards the exterior of the domain.

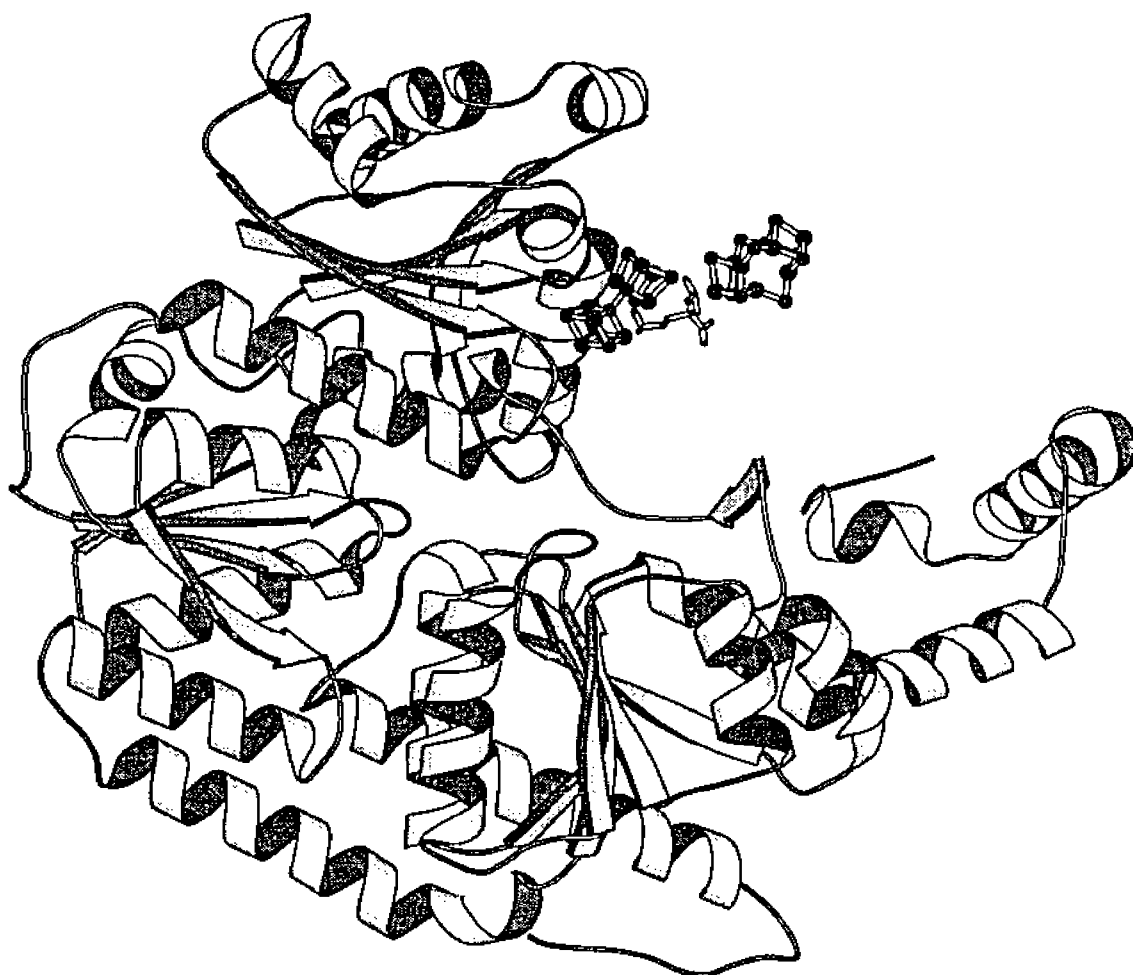


Figure 2.3-3. Molscript drawing of the β -subunit of MoFe. The P-cluster pair is shown at the top of the molecule to the right of domain I'. Domain II' lies to the left on the bottom and domain III' is to the right. The N-terminal strand that forms part of the sheet in domain I of the α -subunit can be seen on the right above domain III'. Domains II' and III' are spanned by the helix H17[B] at the bottom of the subunit.

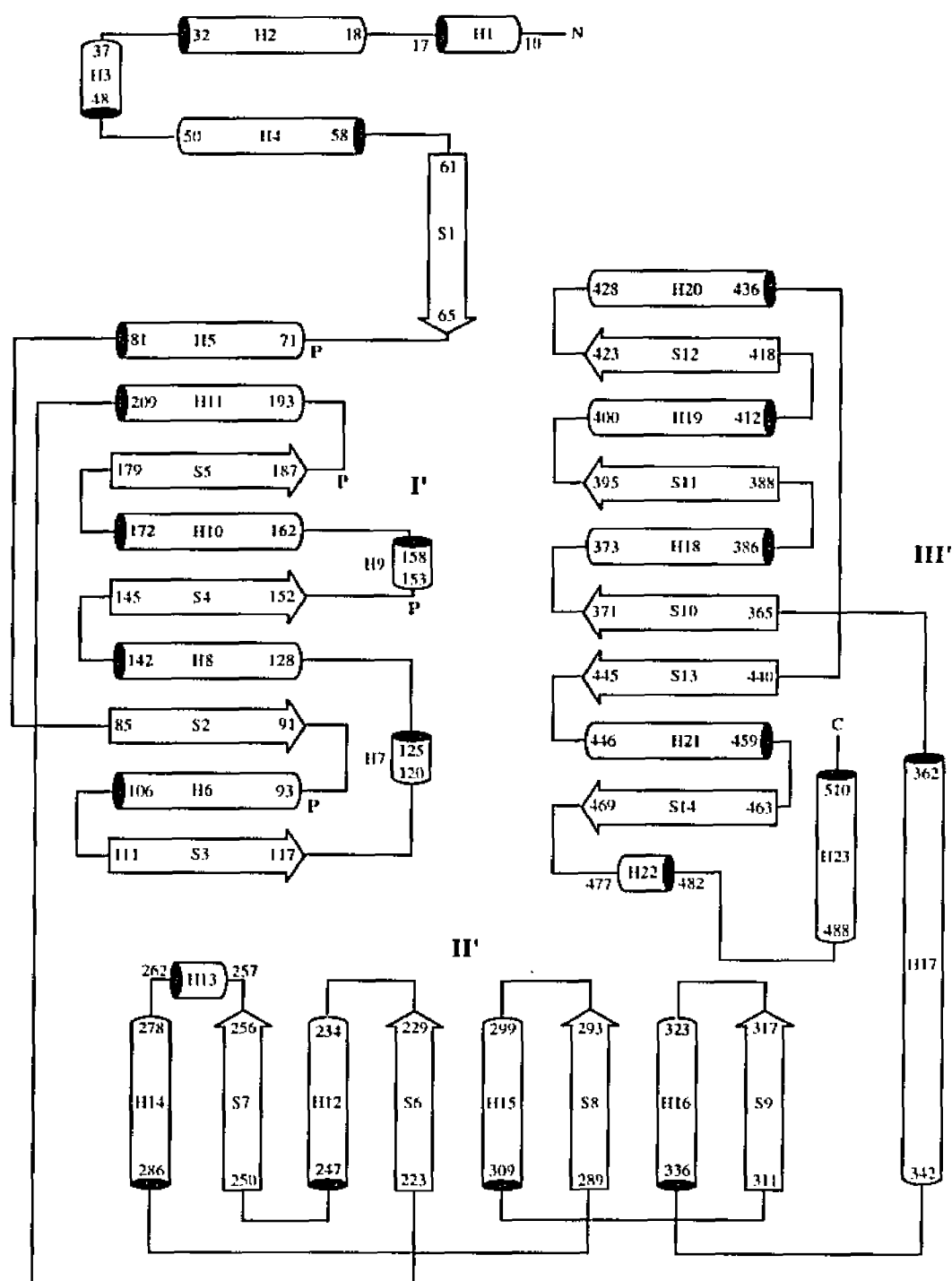


Figure 2.3-4. A schematic representation of the β -subunit of MoFe. The three domains are labeled I', II' and III'. The position of ligands to the P-cluster pair are noted with a P. Note that the β -subunit makes no direct contact with the M-cluster. Within each domain, the shaded helices are located on the same side of the β -sheet in that domain and occupy a more surface exposed region than the unshaded helices.

The α - and β -subunits of the $\alpha\beta$ dimer are related to each other by an approximate two-fold axis that passes through the P-cluster pair (see Figure 2.3-5). As note above, one strand of the β -subunit (from domain I') is bonded antiparallel to the β -sheet in domain I of the α -subunit. The N-terminal helices of the β -subunit can be envisioned as wrapping around the α -subunit like fingers tightly holding this subunit. Additional $\alpha\beta$ -subunit interactions occur with seven salt bridges (between primarily basic residues of the α -subunit and acidic residues of the β -subunit), six hydrophobic patches, four sidechain hydrogen bonds, and many solvent mediated hydrogen bonds.



Figure 2.3-5. Molscript drawing of the $\alpha\beta$ -subunit pair. The view is down the approximate two-fold axis, passing through the P-cluster pair, that relates the α and β subunits. The α -subunit (containing the M-cluster) is shown on the right and is shaded darker than the β -subunit which is on the left. For clarity the helices have been represented as coils.

The complete tetramer is shown below in Figure 2.3-6. It is essentially a dimer of $\alpha\beta$ -subunits related by an approximate two fold non-crystallographic rotation axis. It is

interesting to note that although the two fold axis that relates an $\alpha\beta$ subunit pair is approximately perpendicular (97°) to the tetramer two fold axis, the tetramer does not exhibit the 222 symmetry proposed earlier due to an $\sim 12\text{\AA}$ separation between these rotation axis (Yamane, Weininger et al., 1982). The tetramer interface is predominantly mediated by helical packing interactions between the β -subunits, with additional contributions from salt bridges, hydrogen bonds, hydrophobic interactions, a cation binding site and solvent mediated contacts. For a more detailed discussion of non-solvent mediated interactions, see Kim and Rees (Kim and Rees, 1992).

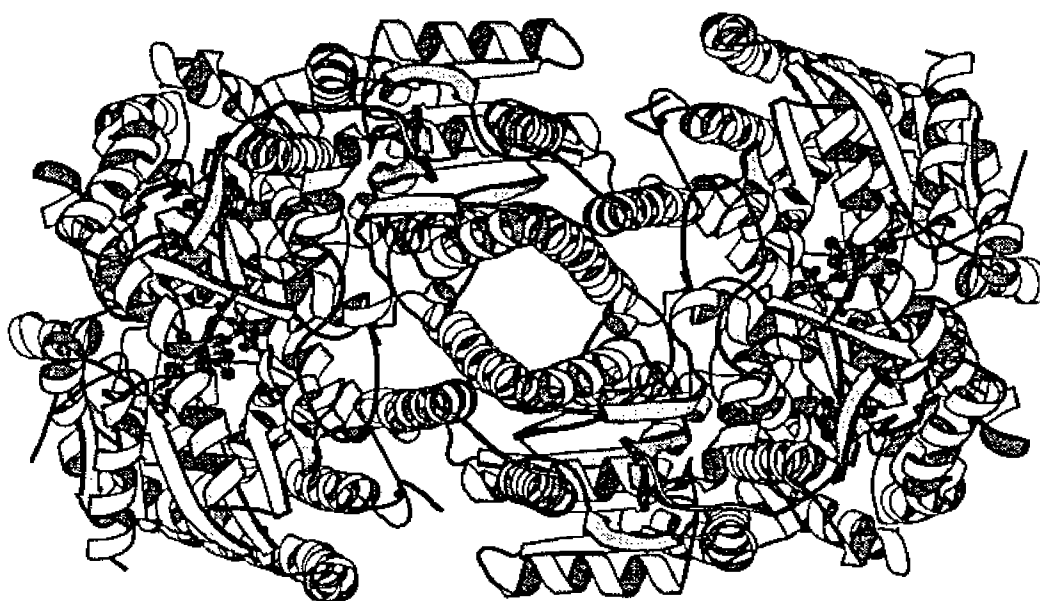


Figure 2.3-6. Molscript drawing of the complete tetramer. The view is down the z-axis along the approximate non-crystallographic two-fold that relates the two $\alpha\beta$ -subunits.

The $\alpha\beta$ -subunit contains 45 proline residues, three of which do not adopt the usual *trans* peptide bond configuration, but instead have been determined to be in a *cis* configuration. These three proline residues are Pro α -449, Pro β -13 and Pro β -472, one of which (Pro α -449) occurs near the FeMo-cofactor binding site in the α -subunit and one of which (Pro β -472) is near the putative binding site in the β -subunit. This means that the entire secondary structure of the MoFe protein contains six *cis*-proline residues. A search of 154 unique structures in the Brookhaven Protein Data Bank by Stewart et al.

(Stewart, Sarkar et al., 1990) found that approximately 6.5% of the prolines are *cis*- rather than *trans*. The occurrence of the *cis*-prolines bond is favored by some sequences, the top four being Tyr-Pro (25%), Ser-Pro (11%), Phe-Pro (10%) and Gly-Pro (8%). In MoFe the *cis*-prolines occur at $\alpha 449$ (Gly-Pro), $\beta 13$ (Tyr-Pro) and at $\beta 472$ (Phe-Pro). The results obtained here for the MoFe protein are consistent with the findings by Stewart et al. both in frequency of occurrence and in residue preference.

Another unusual structural element found in MoFe is the 4.4_{16} or π -helix. This helix had been predicted (Low and Grenville-Wells, 1953), but apparently it has not been observed before, although the catalase from *Penicillium vitale* contains a short helix which is closer to a π -helix than an α -helix (Vainshtein, Melik-Adamyan et al., 1986). To our knowledge MoFe is the first reported occurrence of a π -helix. The π -helix (H3[α]) starts at the P-cluster pair ligand Cys $\alpha 62$ and continues for 12 amino acids, terminating with Pro $\alpha 74$. One of the interesting features of this helix is that initially the conformation is close to that of an α -helix, but after one turn adopts the π -helix conformation. It would be interesting to speculate about the significance with regard to function of such an unusual helix in such close proximity to the P-cluster pair, whose main function is electron transfer of storage during the N_2 reduction cycle (*vide infra*).

The α -subunit of the MoFe-protein also contains one turn of a left-handed α -helix (H19[α]) near one of the FeMo-cofactor ligands (His $\alpha 442$) containing residues Trp $\alpha 444$, Asp $\alpha 445$ and Tyr $\alpha 446$. The left-handed α -helix is an energetically unfavorable structure (Schulz and Schirmer, 1979) and, consequently, only short segments occur (Chou, Nemethy et al., 1990). The presence of this helix in close proximity to the cofactor also provides fertile grounds for speculation as to its ultimate involvement in the mechanism of nitrogen reduction.

Ca⁺² ion

The original model of Kim and Rees contained a divalent cation at the interface between the two $\alpha\beta$ -dimers. Two solvent molecules, the carboxyl oxygens of Glu β -109[B], Asp β -353[D], and Asp β -357[D] along with the carbonyl oxygen of Arg β -108[B], provide an octahedral coordination site for this cation (see Figure 2.3-7). The α -subunit does not contribute a ligand to this site. This cation site was originally modeled as a Mg⁺² or Ca⁺² ion, but the temperature factor suggests that the site is more correctly modeled as Ca⁺². This site is approximately 20Å from the M-cluster and 25Å from the P-cluster pair and presumably has little or no role in nitrogen reduction.

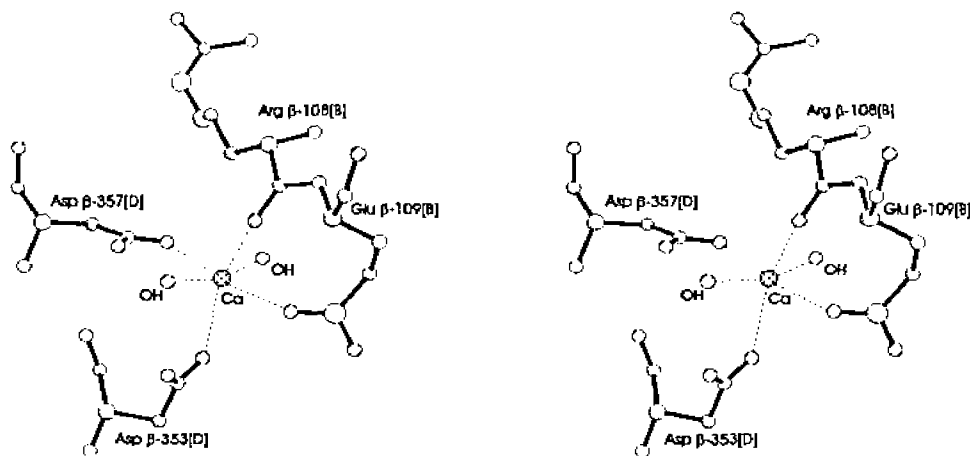


Figure 2.3-7. The cation binding site at the interface of the two $\alpha\beta$ -subunit pairs.

FeMoCo

The FeMo-cofactor (FeMoco) of the MoFe protein has been described in detail previously (Kim and Rees, 1992). Briefly, the cofactor consists of two 4M-3S (where M=transition metal) clusters (derived from a 4M-4S cluster by removing a sulfur atom) bridged by three sulfur atoms (see Figure 2.3-8). One end of the cofactor is composed of a 1Mo:3Fe:3S cubane cluster fragment and the other, a 4Fe:3S cubane cluster fragment. Clusters of the type 4Fe:3S have been previously described (Johansson and Lipscomb, 1958; Chou and Dahl, 1971) and the synthesis of this type of cluster has been recently

reported (Cen, MacDonnell et al., 1994). If the three-fold axis of the two cluster fragments are made coincident with each other, then these two fragments are bridged by sulfur atoms bonded to the metals related by the cluster three-fold axis.

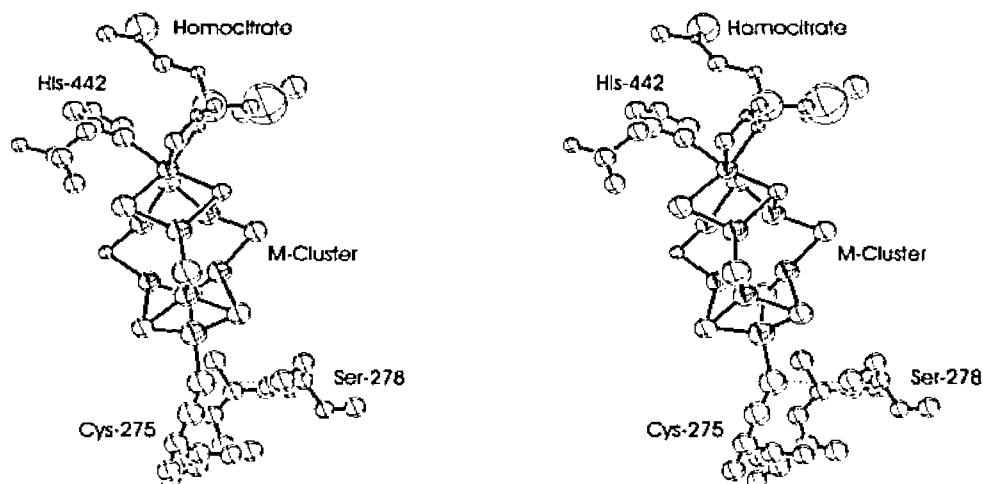


Figure 2.3-8. The MoFe-cofactor and its ligands. The hydrogen bond between the O γ of Ser α 278 and the S γ of Cys α 275 is shown by the dashed line.

Table 2.3-1. Table of selected distances within the MoFe-cofactor. The average distances result from averaging distances in both subunits from the results obtained by refinement in TNT and in XPLOR. Initial distances are the average between the two subunits of 1MIN. The numbers in parenthesis are the standard deviation in the last digit of the averaged value. See Table 2.2.2-4 for a complete list of distances.

From	To	Averaged Distances	Initial Distances
Mo	Fe5	2.76(5)	2.91
	Fe6	2.74(3)	2.97
	Fe7	2.64(5)	2.88
	Fe2	5.04(3)	5.17
	Fe3	5.12(3)	5.31
	Fe4	5.09(5)	5.20
	Fe1	7.10(2)	7.36
Fe2	Fe6	2.57(12)	2.43
Fe3	Fe7	2.62(8)	2.59
Fe4	Fe5	2.56(4)	2.52
Fe1	Fe5	5.06(5)	5.23
	Fe6	5.17(3)	5.15
	Fe7	5.02(4)	5.16
Fe2	Fe5	3.67(6)	3.80
	Fe7	3.55(5)	3.62
Fe3	Fe5	3.72(5)	3.92
	Fe6	3.75(3)	3.90
Fe4	Fe6	3.66(6)	3.77
	Fe7	3.61(5)	3.76

Table 2.3-2. Bond distances within the MoFe-cofactor and to its ligands. The averaged distances are the result of averaging the distances in both subunits from both refinement programs. The numbers in parenthesis are the standard deviation of the last digit for the average. The initial distances are the average distances from 1MIN. See Table 2.2.2-3 for the complete list of bond distances.

From	To	Averaged Distances	Initial Distances
Mo1	S1B	2.41(4)	2.49
	S3B	2.30(4)	2.46
	S4B	2.38(4)	2.44
	Nδ1 α442	2.20(5)	2.25
	O5 CIT	2.31(4)	2.27
	O7 CIT	2.16(6)	2.24
Fe5	S1B	2.27(5)	2.30
	S4B	2.28(7)	2.31
	*S3A	2.24(3)	2.31
Fe6	S1B	2.21(5)	2.29
	S3B	2.23(3)	2.29
	*S2B	2.27(3)	2.28
Fe7	S3B	2.24(2)	2.32
	S4B	2.31(9)	2.36
	*S5A	2.19(2)	1.81
Fe1	S1A	2.25(6)	2.36
	S2A	2.25(5)	2.34
	S4A	2.30(3)	2.34
	Sγ α275	2.27(4)	2.29
Fe2	S1A	2.26(2)	2.30
	S2A	2.23(3)	2.28
	*S2B	2.16(6)	2.27
Fe3	S2A	2.29(2)	2.34
	S4A	2.29(2)	2.39
	*S5A	2.15(4)	1.85
Fe4	S1A	2.34(5)	2.36
	S4A	2.29(5)	2.34
	*S3A	2.14(5)	2.24

The metal-metal distances within the FeMo-cofactor are listed in Table 2.3-1. The average distances of Fe from Mo (including results from both refinement programs) are 2.74(5)Å for the nearest Fe atoms, 5.11(5)Å for the three Fe atoms on the opposite side of the sulfur bridge (i.e., the 4Fe:3S fragment) and 7.14(4)Å to the Fe which is bound to Cys α275. Similar distances are obtained for Fe-Fe distances, i.e., 2.76(4)Å and 5.04(6)Å for the nearest Fe's and the Fe's on the far side of the sulfur bridges respectively. The crystallographically determined Mo-Fe distances compare favorably with the distances determined by EXAFS (Liu, Filipponi et al., 1994) : 2.70Å within the 1Mo:3Fe:3S

fragment and 5.06 Å to the 4Fe:3S fragment. Bond distances for the FeMo-cofactor are shown in Table 2.3-2.

There are only two protein sidechains (Cys α 275 and His α 442) covalently liganded to the FeMo-cofactor; and both of these are contributed by the α -subunit. Both of these cofactor ligands are strictly conserved in all known nitrogenase sequences. The lack of extensive protein ligand binding to the cofactor may account for the ease with which it can be extracted from nitrogenase and could possibly have mechanistic implications by contributing to the overall lability of the cofactor during nitrogen reduction.

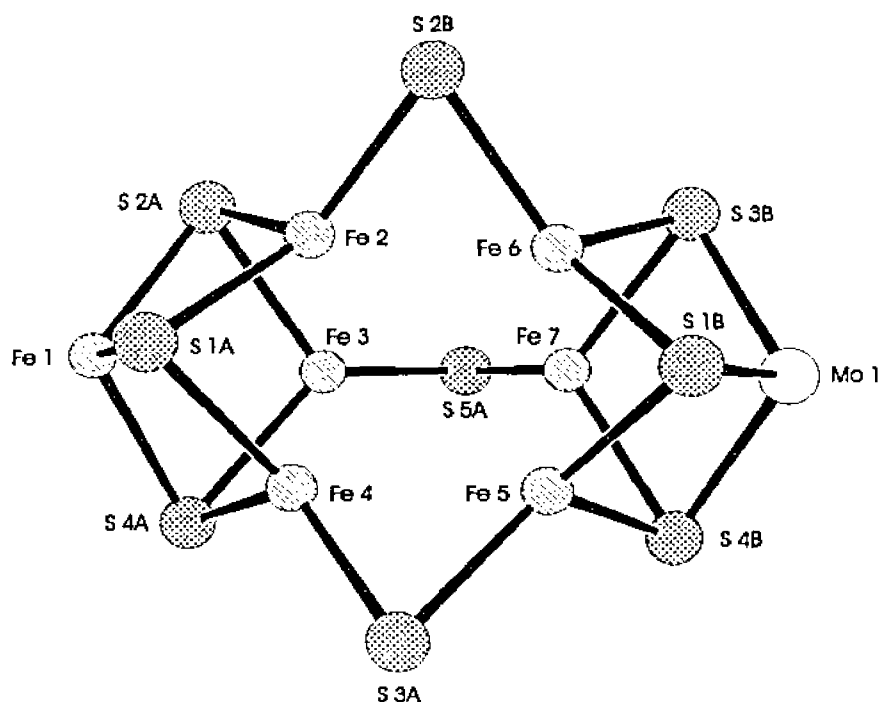


Figure 2.3-9. The M-cluster showing atom labeling scheme.

Homocitrate is bound to the cofactor and has been shown to be essential to the activity of nitrogenase (Hoover, Robertson et al., 1987; Madden, Kindon et al., 1990). Although the homocitrate is covalently bound to the FeMo-cofactor, it forms no covalent bonds to the polypeptide chain. The two octahedral coordination sites of Mo that are not occupied by the sulfur atoms of the cofactor itself or by the only protein ligand to Mo (His α 442) are occupied by two oxygens of the homocitrate, namely O5 and O7 (distances

shown in Table 2.3-3).

FeMo-cofactor environment

There are three residues that are strictly conserved within the FeMo-cofactor environment for all known nitrogenase sequences. These residues are Cys $\alpha 275$, Ser $\alpha 278$ and His $\alpha 442$. The significance of Cys $\alpha 275$ and His $\alpha 442$ is evident by virtue of their function as the only protein ligands to the FeMo-cofactor. However, the importance of Ser $\alpha 278$ may arise from a hydrogen bond formed between its γ -oxygen and the S γ of Cys $\alpha 275$ (see Table 2.3-) which could conceivably serve to stabilize one end of the rather loosely bound MoFe-cofactor. The area around the FeMo-cofactor also contains many highly conserved residues, i.e., Arg $\alpha 96$, Gln $\alpha 191$, His $\alpha 195$, Gly $\alpha 356$, Gly $\alpha 357$, Arg $\alpha 359$, Glu $\alpha 427$ and Glu $\alpha 440$. The glycine residues seem to be necessary in order to provide room for the cofactor. They also form a loop composed of four residues ($\alpha 356$ thru $\alpha 359$) that sits under the cofactor and provides hydrogen bonds to a bridging sulfur (S3 $\alpha 496$) thru their amide nitrogens (see Table 2.3-3). The side chains of the arginine residues $\alpha 359$ and $\alpha 96$ form hydrogen bonds to cluster sulfurs (see Table 2.3-3) and may serve to stabilize the partially reduced intermediates or the cofactor itself. The imidazole sidechain of His $\alpha 195$ is within hydrogen bonding distance of a bridging sulfur (see Table 2.3-3) and may serve to transfer protons or water molecules between the surface of the MoFe protein and the cofactor. The three remaining residues (Gln $\alpha 191$, Glu $\alpha 427$ and Glu $\alpha 440$) are near the homocitrate and interact with it either through solvent mediated contacts, or in the case of Gln $\alpha 191$, directly with it. The side chain of Gln $\alpha 191$ forms a hydrogen bond (see Table 2.3-3) with both a carboxyl oxygen, O4, of the homocitrate and with the amide nitrogen of Gly $\alpha 61$, indicating a possible role in electron transfer.

The hydrophobic residues around the FeMo-cofactor may be divided into two

groups; the more highly conserved residues (Val α 70, Tyr α 229, Ile α 231 and Phe α 381) occupy the area between the P-cluster pair and the FeMo-cofactor, and the non conserved residues (Ile β 355 and Leu α 358) lie between the FeMo-cofactor and the protein surface. The apparent necessity for conserved residues to occur between the two metal centers may have some significance with regard to the electron transfer path between these centers.

Table 2.3-3. Hydrogen bonds to the MoFe-cofactor and its ligands. The values listed are averages that were calculated as they were in the two previous tables.

From		To		Distance (Å)
Residue	Atom	Residue	Atom	
Ser α 278	O γ	Cys α 275	S γ	3.07(22)
Gly α 356	N	CLM α 496	S3A	3.24(10)
Gly α 357	N	CLM α 496	S3A	3.44(6)
Leu α 358	N	CLM α 496	S3A	3.55(3)
Arg α 359	N	CLM α 496	S3A	3.73(6)
Arg α 96	NH2	CLM α 496	S5A	3.09(20)
Arg α 359	NH1	CLM α 496	S4B	3.22(15)
His α 195	NE2	CLM α 496	S2B	3.21(14)
Gln α 191	NE2	CIT α 494	O4	2.77(17)

P-cluster pair

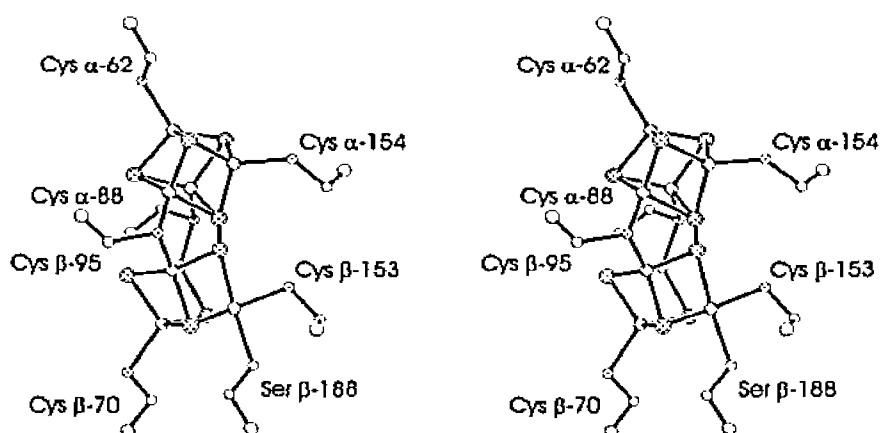


Figure 2.3-10. Stereo view of the P-cluster and its ligands.

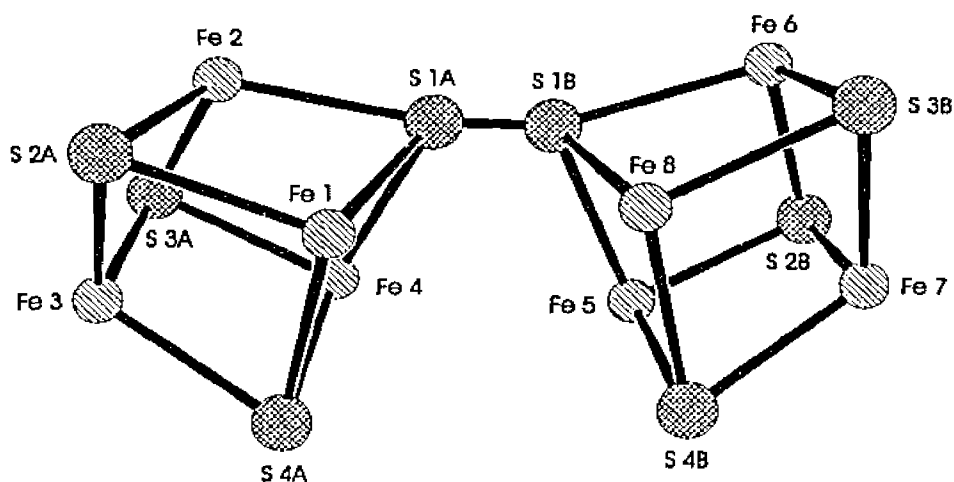


Figure 2.3-11. The P-cluster pair showing the atom labeling scheme.

Table 2.3-4. Selected interatomic distances within the P-cluster pair. Initial distances are the average over both subunits of 1MIN. Average distances are the total average over both subunits and from both refinement packages with the standard deviation of the average in multiples of 0.01Å shown in parentheses. The complete list of distances is shown in Table 2.2.2-6.

From	To	Average Distances	Initial Distances
Fe1	Fe5	4.23(13)	4.33
Fe4	Fe8	4.01(9)	4.03
Fe4	Fe5	3.15(6)	3.11
Fe1	Fe8	2.97(9)	2.98
Fe2	Fe6	5.87(15)	6.23
Fe3	Fe7	7.68(2)	7.79
Fe6	Fe5	3.35(16)	3.40
	Fe7	2.85(7)	2.92
	Fe8	3.14(17)	3.28

The P-cluster pair is composed of two 4Fe:4S cubane clusters bridged by two cysteine thiol ligands and a disulfide bond between two cluster sulfurs (see Figure 2.3-9). In contrast to the lack of protein ligands to the FeMo-cofactor, every metal atom of the P-cluster pair is bonded to an amino acid side chain. The P-cluster pair sits on the approximate two-fold that exists between α and β subunits (see Figure 2.3-5). Both subunits (α and β) contribute ligands to the P-cluster. There are seven strictly conserved ligands, six cysteines (Cys α 62, Cys α 88, Cys α 154, Cys β 70, Cys β 95, Cys β 153) and

one serine (Ser β 188). Selected interatomic distances are given in Table 2.3-4. Bond distances are given in Table 2.3-5.

Table 2.3-5. Table of bond distances within the P-cluster pair and to its ligands. The average distances are calculated for both clusters in the tetramer and include TNT and XPLOR refinement results. Standard deviations are shown in parentheses. The initial distances are averages between from 1MIN. A complete listed of the distances is shown in Table 2.2.2-5.

From	To	Average Distances	Initial Distances
Fe1	S1A	2.28(3)	2.36
	S2A	2.25(9)	2.33
	S4A	2.33(4)	2.39
	S γ β 95	2.24(1)	2.12
Fe2	S1A	2.15(7)	2.05
	S2A	2.26(2)	2.36
	S3A	2.29(5)	2.33
	S γ α 159	2.31(2)	2.37
Fe3	S2A	2.30(5)	2.36
	S3A	2.33(4)	2.38
	S4A	2.30(1)	2.41
	S γ α 62	2.27(1)	2.41
Fe4	S1A	2.26(3)	2.30
	S3A	2.37(1)	2.39
	S4A	2.31(3)	2.34
	S γ α 80	2.26(3)	2.32
Fe5	S1B	2.36(12)	2.36
	S2B	2.27(3)	2.36
	S4B	2.26(4)	2.34
	S γ α 88	2.38(11)	1.98
Fe6	S1B	2.30(12)	2.27
	S2B	2.26(3)	2.29
	S3B	2.36(4)	2.39
	S γ β 153	2.26(3)	2.57
	O γ β 188	2.26(7)	2.17
Fe7	S2B	2.32(2)	2.36
	S3B	2.39(3)	2.37
	S4B	2.30(3)	2.36
	S γ β 70	2.39(1)	2.40
Fe8	S1B	2.02(17)	2.15
	S3B	2.27(4)	2.34
	S4B	2.30(1)	2.33
	S γ β 95	2.30(3)	2.30
S1A	S1B	1.70(26)	2.11

P-cluster pair environment

Hydrophobic residues comprise the majority of the polypeptide environment around the P-cluster pair. Five of these are in the α -subunit (Tyr α 64, Pro α 85, Tyr α 91,

Pro α 155 and Phe α 186), while the β -subunit contributes five more (Pro β 72, Tyr β 98, Phe β 99, Met β 154 and Phe β 189). One of the residues from the β -subunit (Tyr β 98) may serve a role in the transfer of electrons between the P-cluster pair and the MoFe-cofactor by virtue of a solvent mediated contact with the homocitrate. These hydrophobic residues (with the exception of Pro α 85) are not strictly conserved. There is little conservation of the hydrophilic environment around the P-cluster pair among different nitrogenase sequences. The hydrophilic environment is made up of Ser α 92, Ser α 152, Glu α 153, Glu α 184, Ser β 92, Glu β 93 and Thr β 152, with only the latter residues two being strictly conserved. As with the MoFe-cofactor, there are some conserved glycine residues (α 87, α 185 and β 94) that may serve to form a pocket for the cluster.

Solvent Shell

At present, the solvent shell of the MoFe model contains 625 solvent oxygen atoms. The average temperature factor for these solvent atoms is 24.16\AA^2 when refined with TNT (Tronrud, Ten Eyck et al., 1987) and 20.35\AA^2 when refined with XPLOR (Brünger, Kuriyan et al., 1987). These solvent oxygens have been assigned to the subunits where they form the majority of their hydrogen bonds, which results in 137 solvent oxygens in the α -subunit[A], 149 in α -subunit[C], 189 in β -subunit[B] and 150 solvents in β -subunit[D]. Although the majority of the solvent is located at the surface of the tetramer, a significant portion of the solvent is found at the subunit interfaces and in cavities within the tetramer. Six significant cavities can be found in MoFe. A cavity associated with the FeMo-cofactor can be found in each α -subunit, the polypeptide pocket where the FeMo-cofactor is bound. These two cavities were located with the program Voidoo (Kleywegt and Jones, 1994) and have an approximate volume of 550\AA^3 each. A list of the amino acid residues that line these cavities is shown in Table 2.3-6 along with the 19 solvent molecules found in that cavity.

Table 2.3-6. A list of the amino acid residues that line the polypeptide pocket of the α -subunit where the FeMo-cofactor is bound. The residue name and number as well as the chain for the lining of the cavity is shown.

α -Subunit [A]						α -Subunit [C]					
Lining			Contents			Lining			Contents		
ILE	A	59	SOL	A	626	ILE	C	59	SOL	B	710
ARG	A	60	SOL	A	633	ARG	C	60	SOL	C	620
GLY	A	61	SOL	A	634	GLY	C	61	SOL	C	623
CYS	A	62	SOL	A	642	CYS	C	62	SOL	C	625
TYR	A	64	SOL	A	643	TYR	C	64	SOL	C	628
ALA	A	65	SOL	A	656	LYS	C	68	SOL	C	629
LYS	A	68	SOL	A	657	GLY	C	69	SOL	C	644
GLY	A	69	SOL	A	658	VAL	C	70	SOL	C	648
GLY	A	73	SOL	A	674	GLY	C	73	SOL	C	658
PRO	A	74	SOL	A	693	PRO	C	74	SOL	C	659
MET	A	78	SOL	A	694	MET	C	78	SOL	C	665
ILE	A	79	SOL	A	695	ILE	C	79	SOL	C	667
HIS	A	80	SOL	A	700	HIS	C	80	SOL	C	674
SER	A	82	SOL	A	702	ILE	C	81	SOL	C	681
GLY	A	89	SOL	A	703	SER	C	82	SOL	C	682
TYR	A	91	SOL	A	704	GLY	C	89	SOL	C	683
SER	A	92	SOL	A	710	TYR	C	91	SOL	C	706
ARG	A	93	SOL	A	713	SER	C	92	SOL	C	711
ALA	A	94	SOL	B	686	ARG	C	93	SOL	C	728
GLY	A	95				ALA	C	94			
ARG	A	96				GLY	C	95			
ARG	A	97				ARG	C	96			
ASN	A	98				ASN	C	98			
PHE	A	109				PHE	C	109			
VAL	A	110				VAL	C	110			
THR	A	111				THR	C	111			
MET	A	112				MET	C	112			
ASN	A	113				ASN	C	113			
GLN	A	191				PHE	C	114			
GLU	A	380				GLN	C	191			
ILE	A	425				GLU	C	380			
LYS	A	426				ILE	C	425			
HIS	A	442				HIS	C	442			
CIT	A	494				CIT	C	494			
CLM	A	496				CLM	C	496			
CLP	A	498				CLP	C	498			
GLY	B	94				GLY	D	94			
CYS	B	95				CYS	D	95			
ALA	B	97				ALA	D	97			
TYR	B	98				TYR	D	98			
SER	B	101				SER	D	101			
TYR	B	102				TYR	D	102			
ARG	B	105				ARG	D	105			

An interesting feature of this cavity is the observation that the FeMo-cofactor and the P-cluster are accessible to each other through the solvent. The homocitrate is

completely exposed to the solvent in this cavity.

The α -subunit and the β -subunit adopt the same fold, so it is not too surprising that an analogous cavity can be found in the β -subunit in the area of the sequence where the FeMo-cofactor would bind. Much of the space in this cavity that would be occupied by the FeMo-cofactor is instead occupied by the side chains of residues lining the pocket, resulting in a cavity of approximately 425\AA^3 in volume. A list of the residues lining the pocket and the solvent molecules that fill the cavity is given in Table 2.3-7.

Table 2.3-7. Table of the residues lining the cavity in the β -subunit that is analogous to the FeMo-cofactor binding pocket of the α -subunit.

β -Subunit [B]						β -Subunit [D]					
Lining			Contents			Lining			Contents		
GLN	A	90	SOL	A	630	GLN	C	90	SOL	C	622
PRO	B	66	SOL	B	605	PRO	D	66	SOL	D	628
ALA	B	67	SOL	B	612	ALA	D	67	SOL	D	633
LYS	B	68	SOL	B	617	LYS	D	68	SOL	D	635
ALA	B	69	SOL	B	645	ALA	D	69	SOL	D	637
LEU	B	73	SOL	B	667	LEU	D	73	SOL	D	640
TYR	B	102	SOL	B	714	TYR	D	102	SOL	D	653
HIS	B	193	SOL	B	716	HIS	D	193	SOL	D	665
VAL	B	194	SOL	B	731	VAL	D	194	SOL	D	670
PHE	B	230	SOL	B	744	PHE	D	230	SOL	D	681
GLU	B	231	SOL	D	626	GLU	D	231	SOL	D	682
THR	B	232				THR	D	232			
TYR	B	233				TYR	D	233			
ASN	B	236				ASN	D	236			
GLN	B	294				GLN	D	294			
HIS	B	297				HIS	D	297			
ILE	B	318				ILE	D	318			
MET	B	320				MET	D	320			
TRP	B	370				TRP	D	370			
GLY	B	371				GLY	D	371			
ASP	B	372				ASP	D	372			
ASP	B	374				ASP	D	374			
PHE	B	375				PHE	D	375			
ASN	B	397				ASN	D	397			
						LEU	D	427			
						ILE	D	443			
						GLY	D	444			
ASN	B	445				ASN	D	445			
TYR	B	447				TYR	D	447			
ILE	B	469				ILE	D	469			
GLY	B	470				GLY	D	470			
PHE	B	471				PHE	D	471			
PRO	B	472				PRO	D	472			
THR	B	484				THR	D	484			
LEU	B	485				LEU	D	485			

Another area of the MoFe tetramer that contains a significant amount of solvent is the interface between the α - and β -subunits. This interfacial region has the potential to be an area of substrate entry and product release during the catalysis of dinitrogen reduction. Inspection of Table 2.3-8 reveals that these interfacial regions that contain solvent actually contain solvent mediated contacts between three of the four subunits of MoFe. This observation highlights possible factors that may contribute to the overall stability of the MoFe tetramer, i.e., each α -subunit interacts with both of the β -subunits. Another interesting feature of these regions is that they make contact with many of the residues that line the cavity encompassing the FeMo-cofactor and the external surface of the tetramer. This underscores the potential role of the interfacial regions in the transport of substrates and products.

Table 2.3-8. The subunit interface regions that contain significant amounts of solvent.

$\alpha\beta$ -Subunit [ABD]						$\alpha\beta$ -Subunit [CDB]					
Lining			Contents			Lining			Contents		
ALA	A	94	SOL	A	628	ALA	C	94	SOL	B	618
GLY	A	95	SOL	A	652	GLY	C	95	SOL	B	693
ARG	A	96	SOL	A	660	ARG	C	96	SOL	B	703
ARG	A	97	SOL	A	665	ARG	C	97	SOL	C	669
ASN	A	98	SOL	A	688	ASN	C	98	SOL	C	686
TYR	A	99	SOL	A	708	TYR	C	99	SOL	C	705
TYR	A	100	SOL	B	648	TYR	C	100	SOL	C	712
ILE	A	101	SOL	B	685	ILE	C	101			
GLY	A	102	SOL	D	713	GLY	C	102			
THR	A	103	SOL	D	731						
THR	A	104				THR	C	104			
VAL	A	110				VAL	C	110			
THR	A	111				THR	C	111			
SER	A	443				SER	C	443			
TYR	A	446				TYR	C	446			
PHE	B	450				PHE	D	450			
ARG	B	453				ARG	D	453			
GLN	D	513				GLN	B	513			
ASP	D	516				ASP	B	516			
TYR	D	517				TYR	B	517			
ASN	D	518				ASN	B	518			
HIS	D	519				HIS	B	519			
ASP	D	520				ASP	B	520			
LEU	D	521				LEU	B	521			
VAL	D	522				VAL	B	522			
ARG	D	523				ARG	B	523			

There is one other channel from the surface of the tetramer into the polypeptide pocket where the FeMo-cofactor is bound. This channel is on the opposite end of the FeMo-cofactor from the homocitrate and the $\alpha\beta$ -subunit interface. It is completely contained within the α -subunit and is formed by residues Ser α -192, His α -195, His α -196, Arg α -277, Tyr α -281 and His α -383. Although this channel is narrow and may only be large enough in diameter to pass one water molecule at a time, it may provide a pathway for the transfer of substrate, products or protons depending on the conformation of the α -subunit during catalysis.

References

- Brünger, A. T., Kuriyan, J. and Karplus, M. (1987). "Crystallographic R Factor Refinement by Molecular Dynamics." *Science* **235**, 458-460.
- Cen, W., MacDonnell, F. M., Scott, M. J. and Holm, R. H. (1994). "Heterometal Clusters Containing the Cuboidal Fe_4S_3 Fragment: Synthesis, Electron Distribution, and Reactions." *Inorg. Chem.* **33**, 5809-5818.
- Chou, C. T. W. and Dahl, F. (1971). *Inorg. Chem.* **16**, 3245-3251.
- Chou, K. C., Nemethy, G. and Scheraga, H. A. (1990). "Energetic of Interactions of Regular Structural Elements in Proteins." *Acc. Chem. Res.* **23**, 134-141.
- Hoover, T. R., Robertson, A. D., Cerny, R. L., Hayes, R. N., Imperial, J., Shah, V. K. and Ludden, P. W. (1987). "Identification of the V-Factor Needed for Synthesis of the Iron-Molybdenum Cofactor of Nitrogenase as Homocitrate." *Nature* **329**, 855-857.
- Johansson, G. and Lipscomb, W. N. (1958). *Acta. Cryst.* **11**, 594-598.
- Kim, J. and Rees, D. C. (1992). "Crystallographic structure and functional implications of the nitrogenase molybdenum-iron protein from *Azotobacter vinelandii*." *Nature* **360**, 553-560.
- Kim, J. and Rees, D. C. (1992). "Structural Models for the Metal Center in the Nitrogenase Molybdenum-Iron Protein." *Science* **257**, 1677-1682.
- Kleywegt, G. J. and Jones, T. A. (1994). *Voidoo*. Uppsala; pp.
- Liu, H. I., Filipponi, A., Gavini, N., Burgess, B. K., Hedman, B., Di Cicco, A., Natoli, C. R. and Hodgson, K. O. (1994). "EXAFS Studies of FeMo-Cofactor and MoFe Protein: Direct Evidence for the Long-Range Mo-Fe-Fe Interaction and Cyanide Binding to the Mo in FeMo-Cofactor." *J. Am. Chem. Soc.* **116**, 2418-2423.
- Low, B. W. and Grenville-Wells, H. J. (1953). "Generalized Mathematical Relationships for Polypeptide Chain Helices. The Coordinates of the π Helix." *Proc. Natl. Acad. Sci.* **39**, 785-801.

Madden, M. S., Kindon, N. D., Ludden, P. W. and Shah, V. K. (1990). "Diastereomer-dependent substrate reduction properties of a dinitrogenase containing 1-fluorohomocitrate in the iron-molybdenum cofactor." *Proc. Natl. Acad. Sci.* **87**, 6517-6521.

Schulz, G. E. and Schirmer, R. H. (1979). Principles of Protein Structure. In (Editor), Springer Verlag, New York.

Stewart, D. E., Sarkar, A. and Wampler, J. E. (1990). "Occurrence and Role of Cis Peptide Bonds in Protein Structures." *J. Mol. Biol.* **214**, 253-260.

Tronrud, D. E., Ten Eyck, L. F. and Matthews, B. W. (1987). "An Efficient General-Purpose Least-Squares Refinement Program for Macromolecular Structures." *Acta Crystallographica* **A23**, 489-501.

Vainshtein, B. K., Melik-Adamyany, W. R., Barynin, V. V., Vagin, A. A., Grebenko, A. I., Borisov, V. V., Bartels, K. S., Fita, I. and Rossmann, M. G. (1986). "Three-dimensional Structure of Catalase from *Penicillium vitale* at 2.0Å Resolution." *J. Mol. Biol.* **188**, 49-61.

Yamane, T., Weininger, M. S., Mortenson, L. E. and Rossman, M. G. (1982). "Molecular Symmetry of the MoFe Protein of Nitrogenase." *J. Biol. Chem.* **257**, 1221-1223.

2.4 Implications for Nitrogen Reduction

Solvent entry and exit

Access to the FeMo-cofactor from the surface of the protein is required for both the transfer of substrates into the active site and for exit of products after reduction. Regions of the protein that could accommodate substrate/product might also be expected to transport water. There appears to be only one channel in the surface of the α -subunit that permanently contains water. This channel is located where the three domains of the α -subunit come together and is made up of residues $\alpha 45$ - $\alpha 52$, $\alpha 185$ - $\alpha 203$, $\alpha 274$ - $\alpha 285$, $\alpha 354$ - $\alpha 362$ and $\alpha 377$ - $\alpha 389$ (see Figure 2.4-1). This channel funnels into a small passage that is lined by residues Ser α -192, His α -195, His α -196, Arg α -277, Tyr α -281 and His α -383 and allows a nearly direct route to the FeMo-cofactor from the exterior. Although this channel does not appear to be large enough to accommodate the diffusion of substrate/product or H_3O^+ into or out of the active site, it should be emphasized that this structure does not represent the catalytically active form of the enzyme, and conformational changes during catalysis might open and close this channel. A similar mechanism for substrate/product transport is observed in hemoglobin (Case and Karplus, 1979).

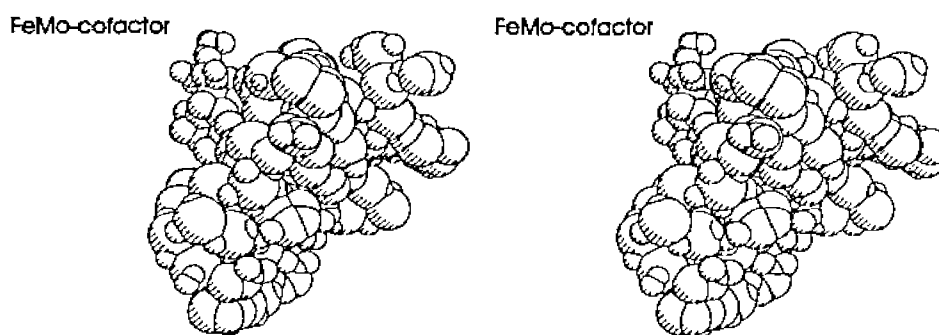


Figure 2.4-1. Stereo view of a space filling model of a cleft in the protein surface (made up of residues $\alpha 45$ - $\alpha 52$, $\alpha 185$ - $\alpha 203$, $\alpha 274$ - $\alpha 285$, $\alpha 354$ - $\alpha 362$ and $\alpha 377$ - $\alpha 389$) that funnels into a solvent filled channel (lined by residues Ser α -192, His α -195, His α -196, Arg α -277, Tyr α -281 and His α -383) leading to the FeMo-cofactor. The FeMo-cofactor is shown at the top left on the side of the protein surface is shown in the lower right, illustrating how close the FeMo-cofactor is to the protein surface and the general topology of the immediate area.

Another possible area where substrate may enter (or product be released from) the active site during catalysis as a result of conformational changes occurs at the interface of domains II and III. This shallow cleft is made up of residues as shown in Figure 2.4-2. This cleft does not appear to taper to a channel that is filled with solvent like the channel discussed above. However, residues making up this cleft appear to be in close proximity with the solvent in the FeMo-cofactor binding pocket. It is reasonable to postulate that any conformational change associated with catalysis could result in enough movement between the three domains to allow passages to appear that would be large enough to allow transport of substrate or product.

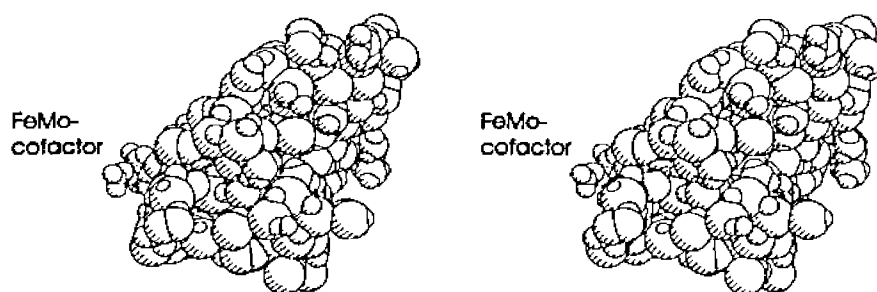


Figure 2.4-2. Stereo view of a space filling model of the cleft in the surface of the α -subunit formed by $\alpha 260$ - $\alpha 288$, $\alpha 293$ - $\alpha 315$, $\alpha 350$ - $\alpha 368$ and $\alpha 378$ - $\alpha 394$.

As discussed in section 2.3, there is a cavity or channel at the α - and β -subunit interface that is filled with solvent (see Table 2.3-8). This channel extends into the polypeptide binding pocket of the FeMo-cofactor. This channel is sufficiently large to allow diffusion of solvent through the channel during the catalytic process. Considering the fact that the β -subunit donates one strand to the β -sheet in domain I of the α -subunit (see Figure 2.3-2), one might expect that conformational changes generated in one subunit may be transmitted to the other. Furthermore, the combination of these may result in the formation of substantial channels for the transport of reactants/substrates/products.

Proton Transfer

Another prerequisite for the reduction of dinitrogen is a steady supply of protons. These protons could be supplied either directly by the free diffusion of H_3O^+ into the active site, or they could be supplied through a mechanism analogous to the one that is operative in the reduction of the quinones of the bacterial photosynthetic reaction center. Reaction center uses a bucket brigade of sorts to supply these protons (Feher, Allen et al., 1986). This bucket brigade type mechanism could involve a patch of His residues near the surface, including His α -196 and His α -383 or His α -274, His α -362 and His α -451. Recall that His α -196 and His α -383 are part of a solvent filled channel in the surface of the α -subunit. The side chains of Asp and Glu residues or water molecules could also be involved in this type of proton transport. The probability is that there are multiple pathways available for the protons needed for reduction to reach the substrates. As discussed in section 2.1, all mutations of His α -195 produce *Nif*⁻ mutants, but MoFe-protein in these proteins is still capable of reducing alternative substrates (Scott, Dean et al., 1992). Many of the histidines listed above were investigated by Thomann et al. during their studies of nitrogen ligation of the FeMo-cofactor (Thomann, Bernardo et al., 1991) and they found there was no significant loss of activity, which points to the availability of several proton transfer pathways. The side chains of Glu α -427 and Glu α -440 provide another possible pathway for proton transfer to the homocitrate. The carboxyl oxygen of Glu α -440 establishes a solvent mediated contact with the homocitrate (Glu α -440 Oe2 – SOL A662 OH = 2.7Å -- CIT α -494 O1 = 3.03Å), but when this residue is mutated to Gly there is no change in the catalytic activity of MoFe-protein (Newton and Dean, 1993). Therefore, it is likely that these two residues are not involved in catalysis. As discussed by Kim and Rees, protons could be transferred to the FeMo-cofactor through Arg α -96 or Arg α -359 if a strong base is generated during the catalytic cycle (Kim and Rees, 1992). Indeed, these two residues may be involved, because mutation to Gln of

Arg α -96 results in a phenotype that still grows diazotrophically but at a slower rate, and a *Nif*⁻ phenotype is the result of the same mutation to Arg α -359 (Newton and Dean, 1993).

Electron Transfer to the FeMo-cofactor

In order for catalysis to occur, electrons must also be delivered to the FeMo-cofactor, in addition to the protons that are required. At this time, there have been no reports of experiments that elucidate the path of electron transfer from the P-cluster pair to the FeMo-cofactor. However, the structure provides some insights as to possible electron transfer pathways. Electrons could be passed the $\sim 14\text{\AA}$ from the P-cluster pair to the FeMo-cofactor either by some through space jumps as seen in other systems (Beratan, Onuchic et al., 1990; Jacobs, Mauk et al., 1991; Wuttke, Bjerrum et al., 1992) or through bond paths involving covalent and/or hydrogen bonds (Cowan, Upmacis et al., 1988).

There are perhaps five possible electron transfer pathways that would involve some through space jumps: 1) from Cys β -95 to Tyr β -98 to homocitrate; 2) from Cys α -88 to Arg α -96; 3) Cys α -88 to Gly α -95 to homocitrate; 4) from Cys α -62 to Val α -72; and 5) Cys α -62 to Ala α -65. There are four helices that lie in parallel between the P-cluster pair and the FeMo-cofactor. These helices are comprised of residues α -63 to α -74, α -88 to α -92, α -191 to α -209 and β -93 to β -106, the first two of which contain ligands to the P-cluster pair while the third helix contains a residue that forms a hydrogen bond to the homocitrate.

Experiments have been conducted on Tyr β -98 in order to test whether or not this residue is involved in the intramolecular electron transfer between the P-cluster pair and the FeMo-cofactor (Peters, 1995). Substitutions of Phe or Leu at this position show virtually no change in activity. However, a substitution of His for Tyr at this position results in a MoFe protein with a specific activity that is significantly lower than the wild type protein. This would seem to suggest that electrons may be passed from the P-cluster pair to the FeMo-cofactor through this area between the two cofactors, and demonstrates that

substitution of a potentially charged residue for Tyr β -98 has a profound effect on the intramolecular electron transfer between these two sites.

The structure also reveals a direct connection between the P-cluster pair and the FeMo-cofactor through covalent and hydrogen bonds. A covalent bond starts the connection with the P-cluster pair ligand Cys α -62 to Gly α -61, then proceeds through two hydrogen bonds from Glu α -61 to Gln α -191 (α -61N - α -191O ϵ 1 = 2.52Å) then from Gln α -191 to the homocitrate (α -191N ϵ 2 - α -494O4 = 2.60Å) and finishes with the covalent bonds between the homocitrate and the FeMo-cofactor. Madden et al. found that a single fluorine for hydrogen substitution on the homocitrate resulted in a substantial reduction in catalytic activity (Madden, Paustian et al., 1991). Modeling shows that this substitution disrupts the hydrogen bond from the homocitrate to Gln α -191. Mutations to Gln α -191 produce a variety of results ranging from a *Nif*⁻ phenotype for Lys α -191 (Scott, Dean et al., 1992), a slow growth phenotype for Ser α -191 and virtually no effect for Ala or Pro α -191 (Newton and Dean, 1993). Again these results seem to indicate that although Gln α -191 is involved in the intramolecular electron transfer between the metal centers, there may be multiple electron transfer pathways available.

Conformational changes during electron transfer

The possibility exists that electron transfer during the catalytic cycle could produce conformational changes in the MoFe protein that involve the polypeptide environment around the FeMo-cofactor. It is interesting to note the presence of relatively infrequent (Chou, Nemethy et al., 1990; Stewart, Sarkar et al., 1990) structural elements in MoFe at what might be viewed as critical areas of the polypeptide. Specifically, the sequence contains a *cis*-proline (Pro α -449) and a left-handed α -helix (H19[α] α -443 to α -447, Figure 2.3-2) near one of the only two FeMo-cofactor ligands, His α -442. A *cis-trans* isomerization of Pro α -449 accompanied by an unwinding of H19[α] could result in a substantial alteration of the FeMo-cofactor binding pocket that may help to accommodate

conformational changes induced in the FeMo-cofactor itself, as discussed below. The other relatively rare structural feature is the π -Helix (H3[a]) found in the α -subunit and beginning at the P-cluster pair ligand Cys α -62. This is also an area where the polypeptide chain spans domains I and III, and if this helix were to adopt an α -helical structure during the catalytic cycle, the interaction of the two domains may be affected, possibly opening another channel for the transport of substrate/product molecules.

M-cluster ligands

Another striking feature of the structure is the lack of ligands to the FeMo-cofactor (two ligands), especially compared to the P-cluster pair (seven ligands). It is generally agreed that catalysts require some type of fluxionality in order to be effective. This lack of ligands may contribute to the overall lability or "breathability" of the FeMo-cofactor. Dance has proposed a theoretical model for dinitrogen reduction that requires conformational changes of the FeMo-cofactor (Dance, 1994). This model proposes binding of dinitrogen to a four iron face of the FeMo-cofactor, that has been made accessible to substrate by the movement of two bridging sulfurs (see Figure 2.4-3).

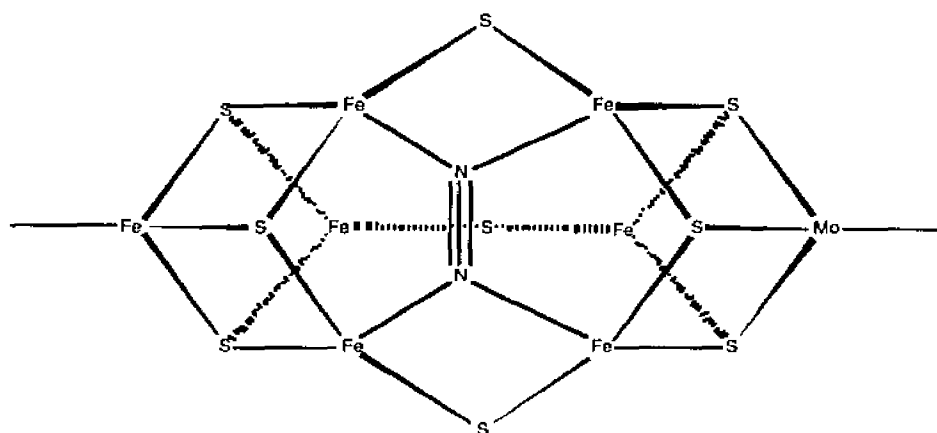


Figure 2.4-3. Nitrogen bound to the face of the FeMo-cofactor at a four iron face with two bridging sulfur atoms folded down and away making the face accessible.

Another dinitrogen binding mode has been suggested by Chan and Rees (Chan, Kim et al., 1993). This proposed binding mode places the N₂ molecule within the central

cavity of the FeMo-cofactor (see Figure 2.4-4). This type of binding would require that the iron to iron bonds between iron atoms bridged by a sulfur be broken and reformed during the catalytic cycle. There is ample precedent for this type of dynamics for iron to iron bonds during catalysis. The crystal structure of $\text{Fe}_2[\mu\text{-S}(\text{CH}_2\text{CH}_3)_2\text{P}(\text{C}_6\text{H}_5)_3(\text{CO})_2]$ reveals an iron to iron bond bridged by two thioethanes with a distance of 2.52\AA (unpublished results, M.W. Day). This Fe-Fe bond has been shown to be labile during catalysis under relatively mild conditions (Aime, Botta et al., 1985). Vahrenkamp et al. have shown that Fe-Fe bond opening is a facile process during the N-N bond cleavage of azoalkanes (Wucherer, Tasi et al., 1989; Hansert, Powell et al., 1991; Hansert, Tasi et al., 1991). This type of binding for dinitrogen is attractive because it could allow all six of the iron atoms to participate in the catalytic process by promoting the rehybridization of the N_2 orbitals before protonation begins and thereby avoiding the high energy intermediates that are generally listed as part of the reaction coordinate.

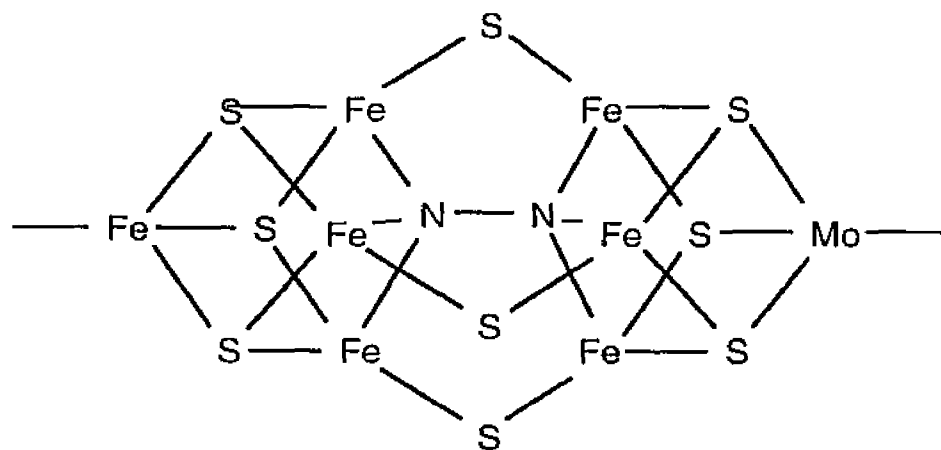


Figure 2.4-4. Dinitrogen binding within the central cavity of the FeMo-cofactor.

References

- Aime, S., Botta, M., Gobetto, R. and Osella, D. (1985). "Activation of the M-CO Bond in Transition Metal Complexes; $\text{Fe}_2(\text{CO})_6(\text{SMe})_2$ and Phosphine Substituted Derivatives as Good Catalysts in Metal-Carbonyl Substitution Reactions." *Organomet.* **4**, 1475-1476.
- Beratan, D. N., Onuchic, J. N., Betts, J. N., Bowler, B. E. and Gray, H. B. (1990). "Electron Tunneling Pathways in Ruthenated Proteins." *J. Amer. Chem. Soc.* **112**, 7915-7921.
- Case, D. A. and Karplus, M. (1979). *J. Mol. Biol.* **132**, 343-368.

- Chan, M. K., Kim, J. and Rees, D. C. (1993). "The Nitrogenase FeMo-Cofactor and P-Cluster Pair: 2.2Å Resolution Structures." *Science* **260**, 792-794.
- Chou, K. C., Nemethy, G. and Scheraga, H. A. (1990). "Energetic of Interactions of Regular Structural Elements in Proteins." *Acc. Chem. Res.* **23**, 134-141.
- Cowan, J. A., Upmacis, R. K., Beratan, D. N., Onuchic, J. N. and Gray, H. B. (1988). "Long-Range Electron Transfer in Myoglobin." *Ann. N. Y. Acad. Sci.* **550**, 68-84.
- Dance, I. G. (1994). "The Binding and Reduction of Dinitrogen at an Fe₄ Face of the FeMo Cluster of Nitrogenase." *Aust. J. Chem.* **47**, 979-990.
- Feher, G., Allen, J. P., Okamura, M. Y. and Rees, D. C. (1986). "Structure and Function of Bacterial Photosynthetic Reaction Centers." *Nature* **339**, 111-116.
- Hansert, B., Powell, A. K. and Vahrenkamp, H. (1991). "Azoalkane- and Nitrene-Bridged Carbonyl Metal Clusters of Iron and Ruthenium." *Chem. Ber.* **124**, 2697-2704.
- Hansert, B., Tasi, M., Tiripicchio, A., Camellini, M. T. and Vahrenkamp, H. (1991). "C-H Cleavage versus N-N Cleavage in m³-h²-Azoalkane Ligands on the M₃(CO)₉ Clusters Iron, Ruthenium, and Osmium." *Organomet.* **10**, 4070-4073.
- Jacobs, B. A., Mauk, M. R., Funk, W. D., MacGillivray, R. T. A., Mauk, A. G. and Gray, H. B. (1991). "Preparation, Characterization, and Intramolecular Electron Transfer in Pentaammineruthenium Histidine-26 Cytochrome *b*₅ Derivatives: Role of the Intervening Medium in Long-Range Donor-Acceptor Electronic Coupling." *J. Amer. Chem. Soc.* **113**, 4390-4394.
- Kim, J. and Rees, D. C. (1992). "Crystallographic structure and functional implications of the nitrogenase molybdenum-iron protein from *Azotobacter vinelandii*." *Nature* **360**, 553-560.
- Madden, M. S., Paustian, T. D., Ludden, P. W. and Shah, V. K. (1991). "Effects of Homocitrate, Homocitrate Lactone, and Fluorohomocitrate on Nitrogenase NifV⁻ Mutants of *Azotobacter vinelandii*." *J. Bacteriol.* **173**, 5403-5405.
- Newton, W. E. and Dean, D. R. (1993). "Role of the Iron-Molybdenum Cofactor Polypeptide Environment in *Azotobacter vinelandii* Molybdenum-Nitrogenase Catalysis." *ACS Symp. Ser.* **535**, 216-230.
- Peters, J. W. (1995). "The Pathway of Electron Transfer within the Nitrogenase Complex." Ph.D. Thesis, Virginic Polytechnic Institute and State University.
- Scott, D. J., Dean, D. R. and Newton, W. E. (1992). "Nitrogenase-catalysed Ethane Production and CO-sensitive Hydrogen Evolution from MoFe Proteins Having Amino Acid Substitutions in an α-Subunit FeMo Cofactor-binding Domain." *J. Biol. Chem.* **267**, 20002-20010.
- Stewart, D. E., Sarkar, A. and Wampler, J. E. (1990). "Occurrence and Role of *Cis* Peptide Bonds in Protein Structures." *J. Mol. Biol.* **214**, 253-260.
- Thomann, H., Bernardo, M., Newton, W. E. and Dean, D. R. (1991). "N Coordination of the FeMo-cofactor requires His-195 of The MoFe Protein a Subunit and is Essential For Biological Nitrogen Fixation." *Proc. Natl. Acad. Sci.* **88**, 6620-6623.
- Wucherer, E. J., Tasi, M., Hansert, B., Powell, A. K., Garland, M.-T., Halet, J.-F., Saillard, J.-Y. and Vahrenkamp, H. (1989). "Intramolecular Conversion of an Azoalkane Ligand to Two Nitrene Ligands on a Triiron Cluster." *Inorg. Chem.* **28**, 3564-3572.

Wuttke, D. S., Bjerrum, M. J., Winkler, J. R. and Gray, H. B. (1992). "Electron Tunneling Pathways in Cytochrome-c." *Science* **256**, 1007-1009.

Chapter 3

Ru(2,2'-bppy)₂(imd)His83 Azurin

from

Pseudomonas aeruginosa

3.1 Introduction to Azurin

Azurin is a relatively small (128 amino acids) electron transfer protein found in several bacteria (Fee, 1975; Lappin, 1981). As the name implies the protein is highly colored, deriving its blue color from a type I, Cu^{+2} (blue copper) center. The azurin from *Pseudomonas aeruginosa* has an approximate molecular weight of 14,000 kD and plays a central role in the organism. Azurin has a relatively high redox potential ($E^{\circ} \approx 300$ -400 mV) and mediates the passage of electrons from cytochrome c_{551} to the cytochrome oxidase/nitrate reductase system. Additionally, azurin has other distinctive spectroscopic properties such as a very intense absorption band in the visible region ($\lambda_{\text{max}} \approx 595$ to 630 nm) and a narrow hyperfine splitting in the EPR spectra ($A_{\parallel} \approx 0.006 \text{ cm}^{-1}$) (Gray and Solomon, 1981). Several xray crystal structures have been reported for the azurins from *Pseudomonas aeruginosa* (Karlsson, Tsai et al., unpublished results; Adman, Stenkamp et al., 1978; Adman and Jensen, 1981; Adman, Canters et al., 1982; Nar, Huber et al., 1992; Nar, Messerschmidt et al., 1992; Nar, Messerschmidt et al., 1992; Tsai, Sjölin et al., 1995), *Alcaligenes denitrificans* (Baker, 1988; Romero, Hoitink et al., 1993; Shepard, Kingston et al., 1993) and *Pseudomonas denitrificans* (Korszun, 1987).

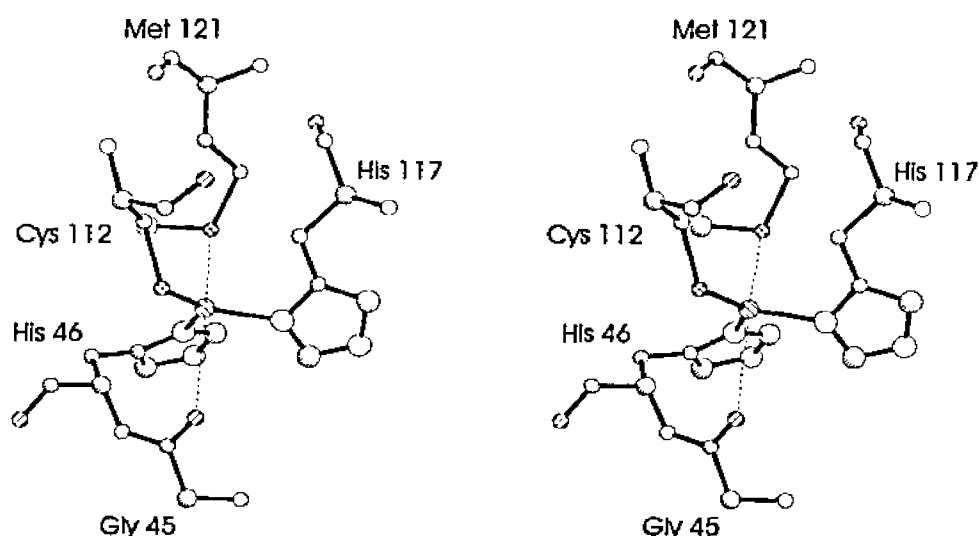


Figure 3.1-1. Stereo view of the Cu site in the azurin from *Pseudomonas aeruginosa*. The three amino acids His 46, Cys 112 and His 117 contribute equatorial ligands and the amino acids Gly-45 and Met 121 supply the axial ligands.

The Cu atom lies approximately 7 Å below the surface of the protein and is liganded by five amino acids. The side chains of His 46, Cys 112 and His 117 occupy equatorial coordination sites with the Sγ of Met 121 and the carbonyl oxygen of Gly 45 in axial positions (see Figure 3.1-1). Therefore, the geometry around copper might be best described as trigonal bipyramidal. The side chain of Met 121 has been suggested to make an important contribution to the fine tuning of the reduction potential (Gray and Malmström, 1983; Pascher, Karlsson et al., 1994). Mutations of Met 121 (Karlsson, Assa et al., 1989), His 46 (Chang, Iverson et al., 1991) and His 117 (den Blaauwen, van de Kamp et al., 1991) result in blue protein solutions. The blue color presumably arises from the *d-d* charge transfer band at 600 nm between the copper atom and the sulfur atom of Cys 112 (Mizoguchi, Di Bilio et al., 1992).

Since azurin is involved in electron transfer, there has been much interest in the electron transfer path through the protein and in the significance of the metal site geometry in the facilitation of the reduction and oxidation of the copper atom. Experiments with Cr⁺³ labeled azurin suggested that there were two sites (His35 and His117) on the surface that participated in electron transfer, one site for cytochrome *c*₅₅₁ and another for nitrate reductase (Farver and Pecht, 1981; Farver, Blatt et al., 1982; Farver, Shahak et al., 1982; Farver and Pecht, 1984). The proposed site around His35 was examined using a crystallographic analysis of the His35Gln and His35Leu mutants by Nar et al. (Nar, Messerschmidt et al., 1992). It had been suggested that deprotonated His35 could compete with the copper ligand, His46, for a hydrogen bond to the carbonyl oxygen of Asn10 (Adman, 1985), but the crystallographic analysis of the His35Gln and His35Leu mutants suggested that this was not the case. The site around His117 was investigated crystallographically in an analysis of the Phe114Ala mutant (Tsai, Sjölin et al., 1995). The area around Phe114 contains the invariant residues Met13, Met44, Phe114, Pro115 and Gly116. Mutation of Phe114 to Ala shows no significant effect on the self-exchange

electron transfer rate, but it did have a slight effect on both the reduction potential (increased by 50 mV) and on the hyperfine splitting (decreased by $4 \times 10^{-4} \text{ cm}^{-1}$) in the EPR spectra.

Temperature jump and stopped flow experiments suggested that the fast electron transfer process in azurin was coupled to a slower process involving a conformational change in reduced azurin (Antonini, Finazzi-Agro et al., 1970; Brunori, Greenwood et al., 1974). NMR experiments showed that the protonation rate of one of the four histidine residues showed an unusual behavior upon pH titration (Ugurbil and Bersohn, 1977). This residue was shown to be His35 and the deprotonation rate was comparable to the slow isomerization rate of azurin (Smith and Smith, 1979) which was also pH dependent (Corin, Bersohn et al., 1983). Crystallographic analysis of wild type azurin demonstrated that the pH dependent behavior was due to a peptide bond flip at His35 (Nar, Messerschmidt et al., 1992).

The unusual spectroscopic properties of the blue copper proteins are generally believed to arise from the protein forcing a certain geometry on the metal site that has functional advantages in that the metal will be liganded in such a way that is optimal for its particular functional role. This idea that the metal fits into a site that is geometrically prepared by the protein has been termed a 'rack' mechanism (Gray and Malmström, 1983). This implies that binding to the metal site is preferential for one cation over another and indeed it has been demonstrated that apo-azurin will take up Cu^{+2} much faster than Ni^{+2} , Co^{+2} and Mn^{+2} (Tennent and McMillin, 1979) and that Cu^{+2} binds more strongly than Zn^{+2} (Engeseth and McMillin, 1986). A number of crystallographic experiments have been done to test this hypothesis. The crystal structures of the apo-azurins from *Pseudomonas aeruginosa* (Nar, Messerschmidt et al., 1992) and *Alcaligenes denitrificans* (Shepard, Kingston et al., 1993) show very little change in the position of the metal ligands, the most significant difference being a slight shifting of His117 toward the copper site in *P.*

aeruginosa. However, Shepard et al. did note a shrinkage in the radius of the metal binding cavity (defined by His46, Cys112 and His117) from 1.36Å in reduced azurin to 1.24Å in oxidized azurin to 1.16Å in apo-azurin. The crystal structure of the zinc containing azurin from *P. aeruginosa* shows movement in the metal ligands by an amount similar to the overall average coordinate error (Nar, Huber et al., 1992).

Numerous experiments using inorganic redox reagents to probe the electron transfer process in azurin have been reported (Antonini, Finazzi-Agro et al., 1970; Lappin, Segal et al., 1979; Farver, Blatt et al., 1982; Gray, 1986; Sykes, 1988). Ruthenium labels have been attached to surface residues to measure the electron transfer rates between the label and the copper site of azurin (Kostic, Margalit et al., 1983; Che, Margalit et al., 1987; Winkler and Gray, 1992). The work described here in Chapter 3 is a crystallographic analysis of the azurin protein from *Pseudomonas aeruginosa* labeled at His83 with $\text{Ru}(2,2'\text{-bppy})_2(\text{imd})_2$ (see Figure 3.1-1).

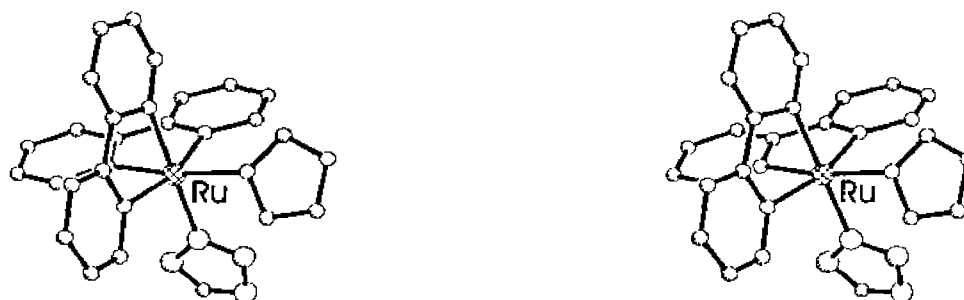


Figure 3.1-1. Stereo view of the $\text{Ru}(2,2'\text{-bppy})_2(\text{imd})_2$ used to label azurin at His83. His83 displaces one imidazole to form a covalent complex.

References

- Adman, E. T. (1985). Structure and Function of small blue copper proteins. Topics in Molecular and Structural Biology: Metalloproteins. 6. Chemie Verlag, Weinheim; pp. 1-42.
- Adman, E. T., Canters, G. W., Hill, H. A. O. and Kitchen, N. A. (1982). "The effect of pH and temperature on the structure of the active site of azurin from *Pseudomonas aeruginosa*." *FEBS Lett.* **143**, 287-292.
- Adman, E. T. and Jensen, L. H. (1981). "Structural Features of Azurin at 2.7Å Resolution." *Israel Journal of Chemistry* **21**, 8-12.
- Adman, E. T., Stenkamp, R. E., Sicker, L. C. and Jensen, L. H. (1978). "A Crystallographic Model for Azurin at 3Å Resolution." *Journal of Molecular Biology* **123**, 35-47.

- Antonini, E., Finazzi-Agro, A., Avigliano, L., Guerrieri, P., Rotilio, G. and Mondovi, B. (1970). "Kinetics of Electron Transfer Between Azurin and Cytochrome c_{551} from *Pseudomonas*." *J. Biol. Chem.* **245**, 4847-4856.
- Baker, E. N. (1988). "Structure of Azurin from *Alcaligenes denitrificans* Refinement at 1.8 Å Resolution and Comparison of the Two Crystallographically Independent Molecules." *Journal of Molecular Biology* **203**, 1071-1095.
- Brunori, M., Greenwood, C. and Wilson, M. T. (1974). "A Temperature-Jump Study of the Reaction Between Azurin and Cytochrome c_{551} from *Pseudomonas aeruginosa*."
- Chang, T. K., Iverson, S. A., Rodrigues, C. G., Kiser, C. K., Lew, A. Y. C., Germanas, J. P. and Richards, J. H. (1991). "Gene Synthesis, Expression, and Mutagenesis of the Blue Copper Proteins Azurin and Plastocyanin." *Proc. Natl. Acad. Sci.* **88**, 1325-1329.
- Che, C. M., Margalit, R., Chiang, H. G. and Gray, H. B. (1987). "Ruthenium Modified Proteins - Reactions of $cis-[Ru(NH_3)_4(H_2O)_2]^{+2}$ and $cis-[Ru(en)_2(H_2O)_2]^{+2}$ with Azurin, Myoglobin and Cytochrome-c." *Inorg. Chem.* **135**, 33-35.
- Corin, A. F., Bersohn, R. and Cole, P. E. (1983). "pH Dependence of the Reduction-Oxidation Reaction of Azurin With Cytochrome c_{551} : Role of Histidine-35 of Azurin in Electron Transfer." *Biochem.* **22**, 2032-2038.
- den Blaauwen, T., van de Kamp, M. and Canters, G. W. (1991). "Type-I and Type-II Copper Sites Obtained by External Addition of Ligands to a His117Gly Azurin Mutant." *J. Amer. Chem. Soc.* **113**, 5050-5052.
- Engeseth, H. R. and McMillin, D. R. (1986). "Studies of Thermally Induced Denaturation of Azurin and Azurin Derivatives by Differential Scanning Calorimetry - Evidence for Copper Selectivity." *Biochem.* **25**, 2448-2455.
- Farver, O., Blatt, Y. and Pecht, I. (1982). "Resolution of Two Distinct Electron Transfer Sites on Azurin." *Biochem.* **21**, 3556-3561.
- Farver, O. and Pecht, I. (1981). Electron transfer processes of blue-copper proteins. Copper Proteins. John Wiley & Sons, Inc. New York; pp. 153-192.
- Farver, O. and Pecht, I. (1984). The reactivity of copper sites in the "blue" copper proteins. Copper Proteins and Copper Enzymes. 1, CRC Press. Boca Raton, FL; pp. 184-214.
- Farver, O., Shahak, Y. and Pecht, I. (1982). "Electron Uptake and Delivery Sites on Plastocyanin in its Reaction with The Photosynthetic Electron Transport System." *Biochem.* **21**, 1885-1890.
- Fee, J. A. (1975). *Struct. Bonding (Berlin)* **23**, 1-60.
- Gray, H. B. (1986). "Long range electron-transfer in blue copper proteins." *Chem. Soc. Rev.* **15**, 17-30.
- Gray, H. B. and Malmström (1983). "On the Relationship between Protein-Forced Ligand Fields and the Properties of Blue Copper Centers." *Comments in Inorganic Chemistry* **2**, 203-209.
- Gray, H. B. and Solomon, E. I. (1981). Electronic Structures of Blue Copper Proteins. Copper Proteins. John Wiley & Sons, Inc. New York; pp. 1-39.

- Karlsson, B. G., Assa, R., Malmström, B. G. and Lundberg, L. G. (1989). "Rack-Induced Bonding in Blue Copper Proteins: Spectroscopic Properties and Reduction Potential of The Azurin Mutant Met121Leu." *FEBS Lett.* **253**, 99-102.
- Karlsson, B. G., Tsai, L.-C., Nar, H., Langer, V. and Sjölin, L. "X-ray structure determination and characterization of the *Pseudomonas aeruginosa* azurin mutant Met121Glu." *unpublished results*
- Korszun, Z. R. (1987). "The Tertiary Structure of Azurin from *Pseudomonas denitrificans* as Determined by Cu Resonant Diffraction Using Synchrotron Radiation." *Journal of Molecular Biology* **196**, 413-419.
- Kostic, N. M., Margalit, R., Che, C. M. and Gray, H. B. (1983). "Kinetics of Long-Distance Ruthenium to Copper Electron Transfer in [Pentaammineruthenium Histidine-83] Azurin." *J. Amer. Chem. Soc.* **105**, 7765-7767.
- Lippin, A. G. (1981). *Met. Ions Biol. Syst.* **13**, 15-71.
- Lippin, A. G., Segal, M. G., Weatherburn, D. C., Henderson, R. A. and Sykes, A. G. (1979). "Kinetic Studies on 1:1 Electron Transfer Reactions Involving Blue Copper Proteins. 3. Protonation Effects, Protein-Complex Association, and Binding Sites in Reactions of *Pseudomonas aeruginosa* azurin with $\text{Co(4,7-DPSphen)}_6^{-3}$, Fe(CN)_6^{-3} , and Co(phen)_3^{+3} (oxidants) and Fe(CN)_6^{-4} (reductant)." *J. Amer. Chem. Soc.* **101**, 2302-2306.
- Mizoguchi, T. J., Di Bilio, A. J., Gray, H. B. and Richards, J. H. (1992). "Blue to Type 2 Binding. Copper (II) and Cobalt (II) Derivatives of a Cys112Asp Mutant of *Pseudomonas aeruginosa* Azurin." *J. Am. Chem. Soc.* **25**, 10076-10078.
- Nar, H., Huber, R., Messerschmidt, A., Filippou, A. C., Barth, M., Jaquinod, M., van de Kamp, M. and Canters, G. W. (1992). "Characterization and crystal structure of zinc azurin, a by-product of heterologous expression in *Escherichia coli* of *Pseudomonas aeruginosa* copper azurin." *European Journal of Biochemistry* **205**, 1123-1129.
- Nar, H., Messerschmidt, A., Huber, R., van de Kamp, M. and Canters, G. W. (1992). "Crystal structure of *Pseudomonas aeruginosa* apo-azurin at 1.85 Å resolution." *FEBS* **306**, 119-124.
- Nar, H., Messerschmidt, A., Huber, R. and van de Kamp, M. (1992). "X-ray Crystal Structure of the Two Site-specific Mutants His35Gln and His35Leu of Azurin from *Pseudomonas aeruginosa*." *Journal of Molecular Biology* **218**, 427-447.
- Pascher, T., Karlsson, B. G., Nordling, M., Malmstrom, B. G. and Vanngard, T. (1994). "Reduction Potentials and Their pH Dependence in Site-Directed-Mutant Forms of Azurin from *Pseudomonas aeruginosa*." *Eur. J. Biochem.* **212**, 289-296.
- Romero, A., Hoitink, C. W. G., Nar, H., Huber, R., Messerschmidt, A. and Canters, G. W. (1993). "X-ray Analysis and Spectroscopic Characterization of M121Q Azurin; A Copper Site Model for Stellacyanin." *J. Mol. Biol.* **229**, 1007-1021.
- Shepard, W. B., Kingston, R. L., Anderson, B. F. and Baker, E. N. (1993). "Structure of Apo-Azurin from *Alcaligenes denitrificans* at 1.8 Å Resolution." *Acta Cryst.* **D49**, 331-343.
- Smith, H. A. O. and Smith, B. E. (1979). "Characteristics of Azurin From *Pseudomonas aeruginosa* via 270 MHz Proton Nuclear Magnetic Resonance Spectroscopy." *J. Inorg. Biochem.* **11**, 79-93.
- Sykes, A. G. (1988). "Electron Transfer in Biological Systems." *Chem. Brit.* 551-554.
- Tennent, D. L. and McMillin, D. R. (1979). *J. Amer. Chem. Soc.* **101**, 2307-2311.

Tsai, L.-C., Sjölin, L., Langer, V., Pascher, T. and Nar, H. (1995). "Structure of the Azurin Mutant Phe114Ala from *Pseudomonas aeruginosa* at 2.6Å Resolution." *Acta Cryst.* **D51**, 168-176.

Ugurbil, K. and Bersohn, R. (1977). "Nuclear Magnetic Resonance Study of Exchangeable and Non-exchangeable Protons in Azurin in *Pseudomonas aeruginosa*." *Biochem.* **16**, 3016-3023.

Winkler, J. R. and Gray, H. B. (1992). "Electron-Transfer in Ruthenium-Modified Proteins." *Chem. Rev.* **92**, 369-379.

3.2 Experimental

Crystal Growth

The hanging drop vapor diffusion method (Weber, 1991) was used for crystallization of the protein-Ru(2,2'-bppy)₂(imd) complex. Two different sets of conditions yielded three crystal forms. One set of conditions yielded light brown crystals with an apparent hexagonal morphology belonging to the trigonal space group $P\bar{3}_121$ with $a=b=76.94\text{\AA}$ and $c=71.29\text{\AA}$ (crystal form I). These crystals grew from a 4 μL drop containing 2 μL of protein solution at approximately 30 mg/ml and 2 μL of well solution (see below) which was equilibrated against a well solution of 2.2M (NH₄)₂SO₄ buffered at pH 5.7 with 100 mM sodium citrate. Crystals appeared in two days. This same set of conditions yielded strongly dichroic crystals of an apparent orthorhombic habit (crystal form II). These crystals were green when viewed perpendicular to two crystal faces but appeared brown when viewed perpendicular to the third crystal face. Although this crystal form diffracted strongly, it was not possible to obtain single crystals, so further work on this form was abandoned. Interestingly, the azurin mutants, H35N and H35L, also grow in a crystal form that exhibits similar dichroism (Nar, Messerschmidt et al., 1992).

A different crystal form (form III) was obtained by equilibrating a 7 μL drop containing 3.5 μL of well solution and 3.5 μL protein solution at approximately 30 mg/ml against a well solution of 30% w/v PEG 4K, 100 mM LiNO₃ and 20 mM CuCl₂ buffered at pH 8.0 with 100 mM Tris. These conditions were similar to those determined by S. Fahim for the crystallization of the azurin Cys112Asp mutant. After one week, crystals belonging to the monoclinic space group $C2$ appeared with cell dimensions $a=100.6\text{\AA}$, $b=35.4\text{\AA}$, $c=74.7\text{\AA}$ and $\beta=106.5^\circ$.

Data Collection

X-ray diffraction data were collected for both crystal forms I and III with a Siemens X-1000 multi-wire area detector using x-rays generated by a Siemens rotating copper

anode, with Cu K_{α} radiation ($\lambda=1.5418\text{\AA}$) selected by a graphite crystal monochrometer. Data for crystal form I were collected at room temperature while data for crystal form III were collected at approximately 100K using the cryogenic techniques of Hope (Hope, 1988; Hope, 1990).

Crystal form I diffracts to a maximum resolution of 2.4\AA . A total of 9737 reflections (10219 possible, 95% complete) were collected from 93280 observations. The merging R-factor was 9.8% and the average $I/\sigma I=8.8$ with a Wilson $B=62.9\text{\AA}^2$.

Crystal forms III diffracts to 2.3\AA resolution. Collection of 38571 observations of 11083 reflections (13540 possible, 82% complete) yielded a data set with merging R-factor of 6.3% with an average $I/\sigma I=16.7$ and an overall B-value calculated from Wilson statistics of 34.3\AA^2 .

Structure Solution

The structure was solved by molecular replacement. The H35Q mutant of *Pseudomonas aeruginosa* solved by Nar et al. was used as the search model (Nar, Messerschmidt et al., 1992). The orientation of the model was determined with the fast rotation function of Crowther and Blow (Crowther and Blow, 1967; Crowther, 1972) using all data between 8 and 3.5\AA with an integration radius of 17\AA . The self-rotation function indicated two molecules related by non-crystallographic symmetry in the asymmetric unit, consistent with our estimation of eight molecules in the unit cell based the ratio of cell volume to molecular weight (Matthews, 1968) where $V_m=2.18\text{\AA}^3/\text{Da}$. The cross rotation function solution contained two peaks which were related to each other by the same angles as the two peaks of the self-rotation function.

The position in the unit cell of one properly oriented molecule was determined by a brute force translation search using data between 8 and 5\AA while searching the xz plane. The position of the other molecule was determined by fixing the position of the first molecule and performing the same brute force translation search between $x=0$ to 1, $y=0$ to

1/2 and $z=0$ to 1. The resulting solution gave a correlation coefficient approximately twice that of the next highest peak. The crystallographic R -factor ($R=\Sigma ||F_o| - |F_c|| / \Sigma |F_o|$) for the rotated and translated model (calculated with the refinement package TNT (Tronrud, Ten Eyck et al., 1987)) was $R=47.9\%$.

Refinement

Difference Fourier electron density maps (Fo-Fc) were calculated and the maps revealed two peaks seven times stronger than the standard deviation of the map near His83. These peaks were located approximately 2.0Å from Nε2 of the imidazole side chain and were approximately in the plane of the ring and therefore were modeled as the ruthenium atom of the label. The remaining ligands of the label were not defined in these maps. The model resulting from rigid body refinement (including the ruthenium of the label) was refined further using the stereochemically restrained least squares module of TNT using all observed data between 7.0 and 2.5Å resolution. Following each round of refinement, the model was evaluated against electron density maps (Fo-Fc and 2Fo-Fc) using the program Tom (Jones, 1978). The model was manually adjusted to fit the electron density map where needed and solvent oxygen atoms were added wherever it was appropriate to do so. After the first round of refinement, two additional copper atoms were modeled into strong difference electron density near the N-terminus. Copper was selected for modeling these peaks because CuCl₂ had been used in the crystallization of the protein. After five rounds of refinement, it became possible to place the bipyridine and imidazole ligands of the ruthenium into electron density. For all subsequent cycles of refinement, stereochemical restraints were applied to the Ru(2,2'-bppy)₂(imd) ligand. (See Appendix 3 for the dictionary of restraints.) Twelve rounds of refinement produced the current model, including 150 solvent oxygen atoms, and has resulted in a crystallographic residual (R -factor) of 17.6%. The rms. deviation of the bond distances and angles from the target values is 0.017Å and 2.70° respectively with the rms. difference in the temperature factors

of adjacent atoms being 3.6\AA^2 . The average temperature factors for the main chain atoms and the side chain atoms are $B=24.76\text{\AA}^2$ and $B=31.65\text{\AA}^2$ respectively for molecule A and $B=22.37\text{\AA}^2$ and $B=28.74\text{\AA}^2$ for molecule B. The average temperature factor for solvent oxygens is $B=42.19\text{\AA}^2$. A plot of the average temperature vs. the residue number is shown in Figure 3.2-1 and representative electron density is shown in Figure 3.2-2.

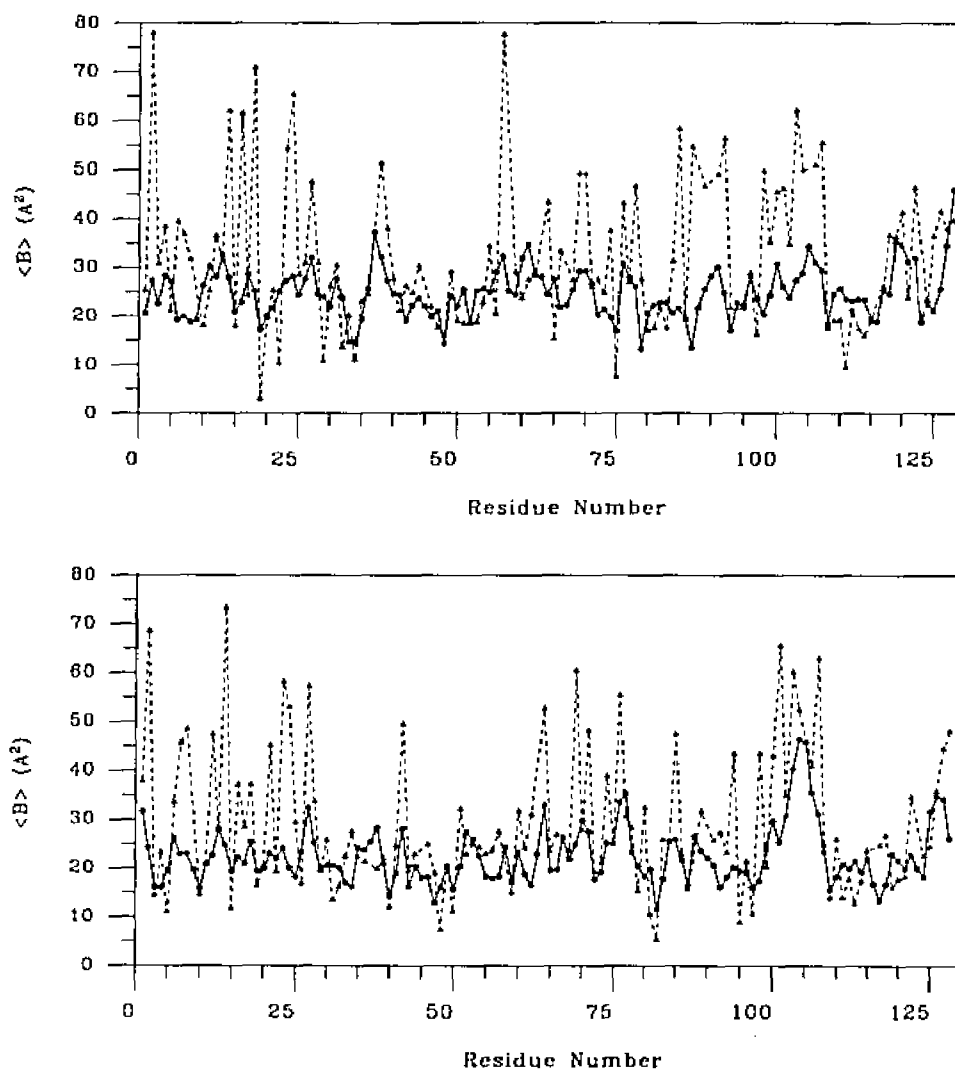


Figure 3.2-1. Plot of average temperature factors vs. residue number. The average values for the main chain atoms are shown with solid lines and the values for the side chain atoms are represented by dashed lines. The top panel shows the plot for molecule A and the bottom panel shows the plot for molecule B.

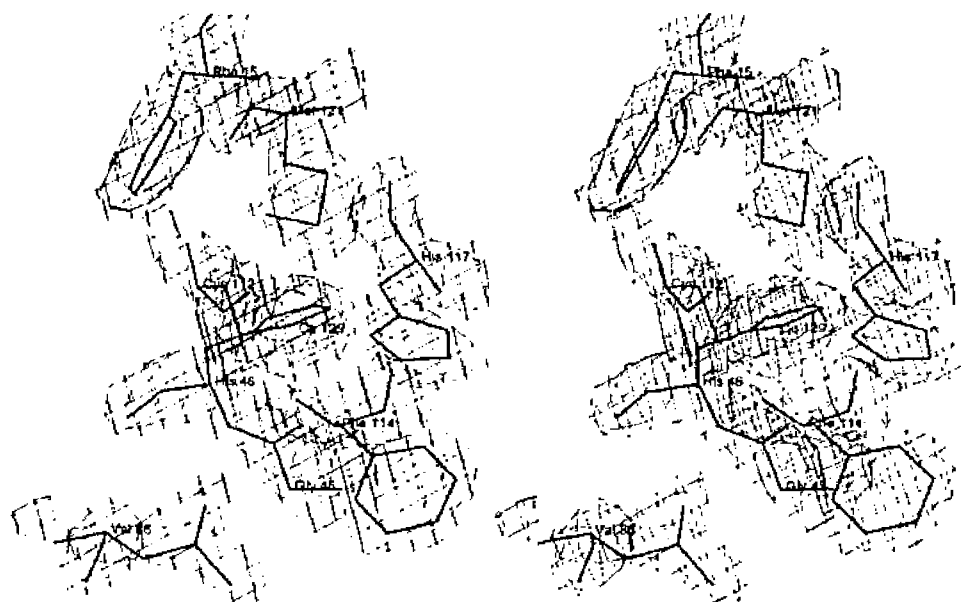


Figure 3.2-2. Stereo view of representative electron density for azurin. The view is for the redox active copper atom environment.

A Ramachadran plot of the (Φ , Ψ) main chain torsion angles is shown in Figure 3.2-3. All of the residues have Φ , Ψ angles in allowed regions.

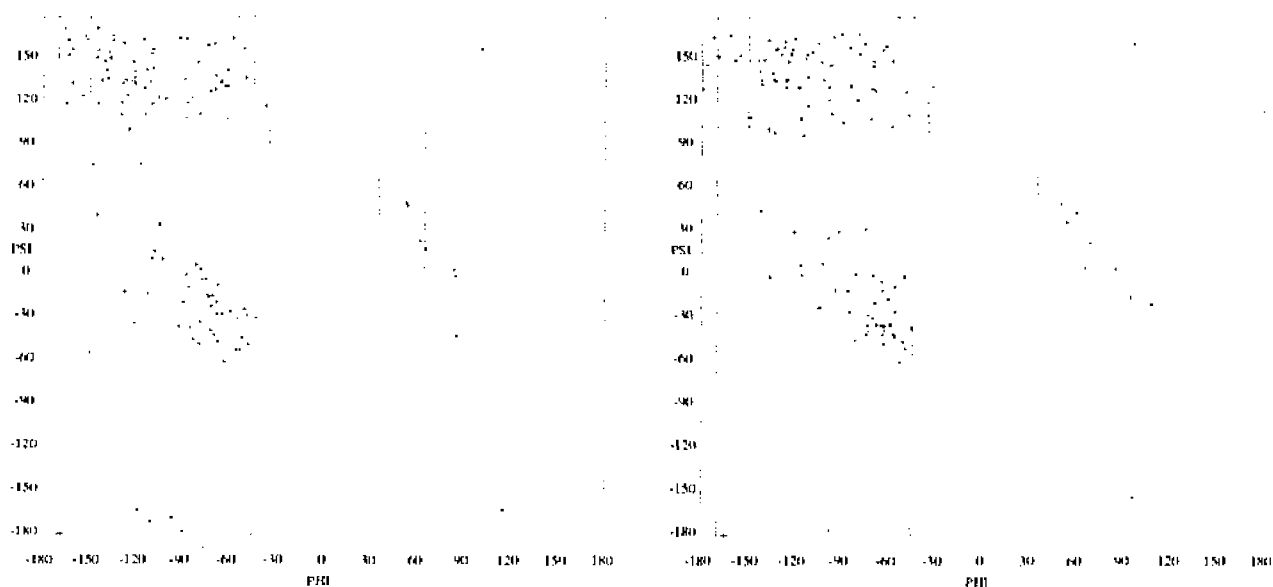


Figure 3.2-3. Ramachadran plot of the (Φ , Ψ) angles for labeled azurin. Residues represented by + are glycines. Molecule A is shown at the left and molecule B is shown to the right.

The average coordinate error can be estimated by the method of Luzzati (Luzzati,

1952), where the dependence of the R-factor on resolution is compared to theoretical values. This analysis is shown in Figure 3.2-4 and results in an estimated average coordinate error of approximately 0.22Å.

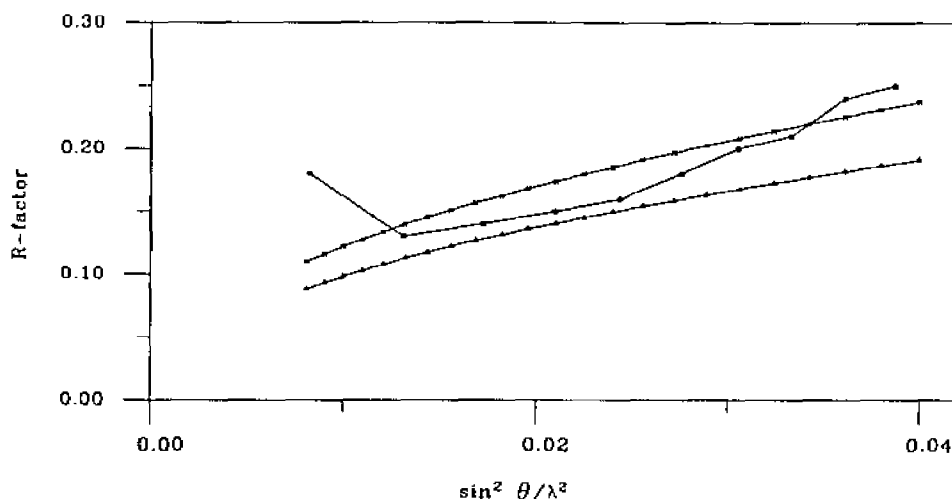


Figure 3.2-4. Plot of the dependence of R-factor on resolution for Ru(2,2'-bipyridine)₂(indole) labeled azurin. Theoretical curves are shown for an estimated error of 0.25Å (top) and 0.20Å (bottom). Only reflections between 7 – 2.5Å resolution were used in the refinement.

References

- Crowther, R. A. (1972). A Method of Positioning a Known Molecule in an Unknown Crystal Structure. The Molecular Replacement Method. Gordon and Breach, New York; pp. 173-178.
- Crowther, R. A. and Blow, D. M. (1967). "A Method of Positioning a Known Molecule in an Unknown Crystal Structure." *Acta Crystallographica* **23**, 544-548.
- Hope, H. (1988). "Cryocrystallography of Biological Macromolecules: a Generally Applicable Method." *Acta Crystallographica* **B44**, 22-26.
- Hope, H. (1990). "Crystallography of Biological Macromolecules At Ultra-Low Temperature." *Annual Review of Biophysics and Biophysical Chemistry* **19**, 107-126.
- Jones, A. T. (1978). "A Graphics Model Building and Refinement System for Macromolecules." *J. Appl. Cryst.* **11**, 268-272.
- Luzzati, V. (1952). "Traitement Statistique des Erreurs dans la Determination des Structures Cristallines." *Acta Cryst.* **5**, 802-810.
- Matthews, B. W. (1968). "Solvent Content of Protein Crystals." *Journal of Molecular Biology* **33**, 491-497.
- Nar, H., Messerschmidt, A., Huber, R. and van de Kamp, M. (1992). "X-ray Crystal Structure of the Two Site-specific Mutants His35Gln and His35Leu of Azurin from *Pseudomonas aeruginosa*." *Journal of Molecular Biology* **218**, 427-447.

Tronrud, D. E., Ten Eyck, L. F. and Matthews, B. W. (1987). "An Efficient General-Purpose Least-Squares Refinement Program for Macromolecular Structures." *Acta Crystallographica* **A23**, 489-501.

Weber, P. C. (1991). "Physical Principles of Protein Crystallization." *Advances in Protein Chemistry* **41**, 1-36.

3.3 Structure Discussion

Overall Fold



Figure 3.3-1. Molscript representation of the azurin from *P. aeruginosa* showing the overall folding topology of a crystallographically unique monomer. This view is the standard view set out by Adman et al. with the copper at the upper (northern) end of the β -barrel and the short α -helix to the right. The copper of crystallization can be seen at the bottom of the molecule and the $\text{Ru}(2,2'\text{-bipyridine})_2(\text{imd})$ can be seen on the back side of the protein.

The overall structure of the labeled azurin from *P. aeruginosa* is very similar to the azurin from *A. denitrificans* and is shown in Figure 3.3-1. An alpha carbon trace of the backbone is shown in Figure 3.3-2. The structure consists of eight β -strands which form a β -barrel and a short stretch of α -helix composed of residues 55 thru 67. There is one

disulfide bond in the molecule formed between the Sγ atoms of Cys3 and Cys26. The protein crystallizes with two crystallographically independent molecules in the asymmetric unit (see Figure 3.3-3). When these two molecules are superimposed onto each other, the rms. deviation of alpha carbon positions is 0.36Å. However, the alpha carbons of the copper ligands (Gly45, His46, Cys112, His 117 and Met121) superimpose on each other with an rms. deviation of 0.16Å. When the labeled azurin is superimposed onto the structure of the azurin from *P. aeruginosa* (Nar, Messerschmidt et al., 1992), the rms. deviation in alpha carbon positions is approximately 0.5Å. When the alpha carbons of the copper ligands are superimposed, the rms. deviation between the two structures is approximately 0.17Å. The deviation in the alpha carbon positions for both of the above superpositions is close to the estimated average coordinate error. Furthermore, the two molecules of the labeled azurin agree with each other to the same degree, they each agree with the structure of the native azurin as determined by Nar et al.. Consequently, this suggests that attachment of the Ru(2,2'-bppy)₂(imd) label has very little or no effect on the overall structure of azurin.

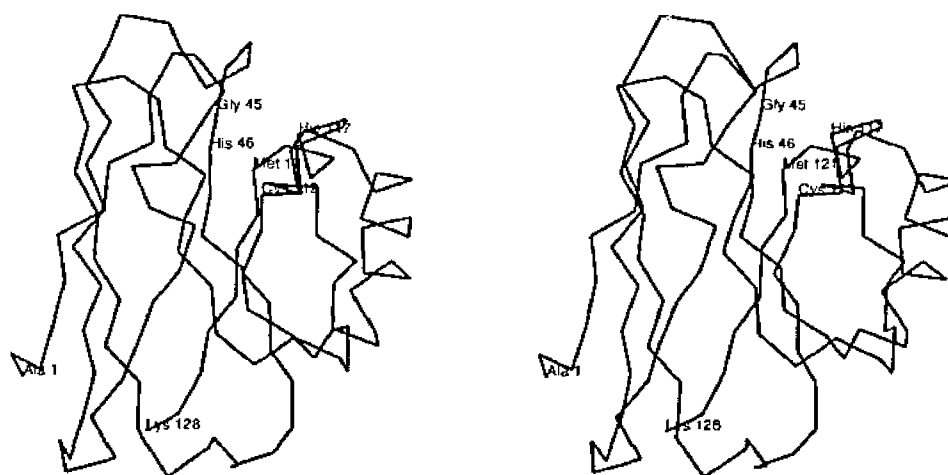


Figure 3.3-2. Stereo view of the alpha carbon trace illustrating the overall fold. This is the standard view with the copper atom at the northern end with the ligands to copper labeled. The N-terminal and C-terminal residues are labeled at the bottom of the molecule.

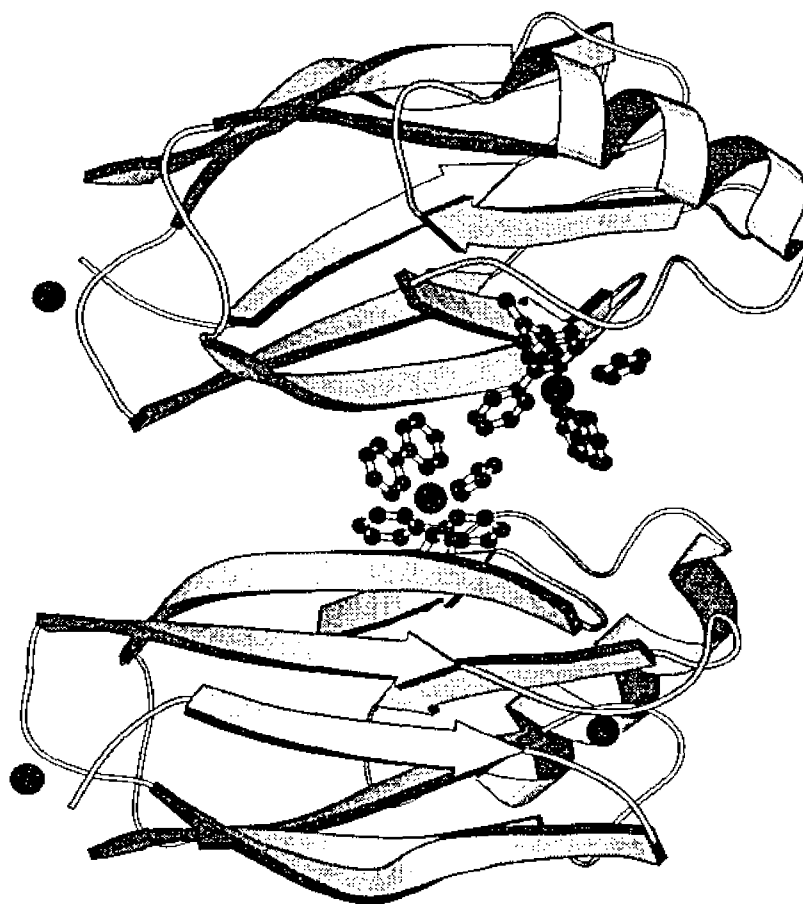


Figure 3.3-3. Molscript drawing of the dimer of crystallization. The copper atom in the active site can be seen inside and at one end of each β -barrel. The $\text{Ru}(2,2'\text{-bipyridine})_2(\text{imd})$ complex can be seen at the interface of the two molecules. The copper of crystallization is visible near the N-terminus of each molecule.

Copper Site

The geometry of the copper site can be described as trigonal bipyramidal and is virtually the same as the copper site in the wild type unlabeled protein (see Figure 3.3-4). The five ligands to copper are His46 N δ 1, Cys112 S γ and His117 N δ 1 in the equatorial plane with Gly45 O and Met121 S δ in the axial positions. A list of bond distances for this structure and for other high resolution structures is given in Table 3.3-1. The distances determined here are very close to those determined for other azurin structures. However, it

should be noted that these distances were restrained during the refinement process (see Appendix 3 for a dictionary of the restraints).

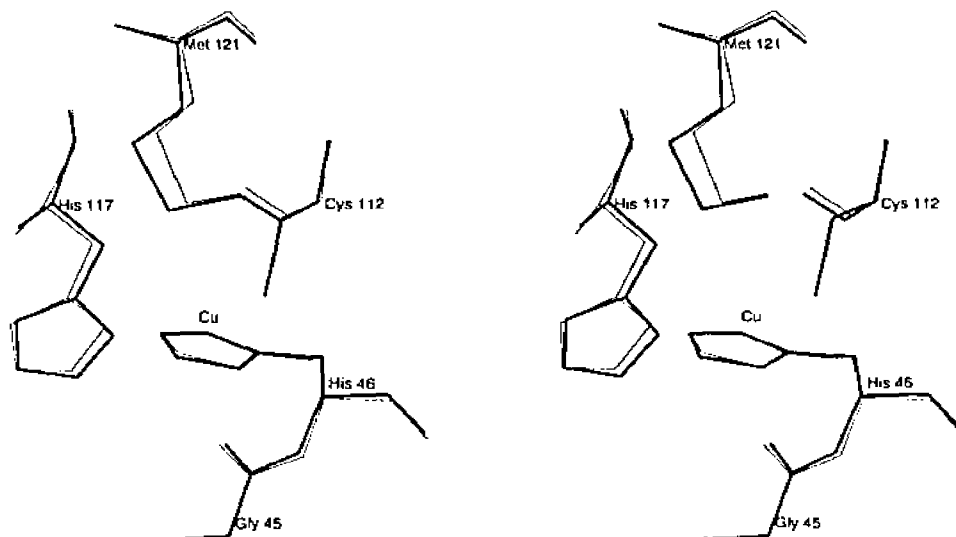


Figure 3.3-4. Stereo view of a superposition of the copper site in the unlabeled azurin on the ruthenium labeled azurin. The thick lines represent the ruthenium labeled azurin and the thin lines represent the structure described by Nar et al. (Nar, Messerschmidt et al., 1992).

Table 3.3-1. Copper ligand distances (Å) for the high resolution azurin crystal structures reported to date. Ru A and Ru B are this structure, H35L and H35Q are the *P. aeruginosa* mutants reported by Nar et al. (Nar, Messerschmidt et al., 1992) and listed in the PDB as 2AZU and 3AZU respectively, AD is the structure of *A. denitrificans* listed in the PDB as 2AZA (Baker, 1988), Zn is the Zinc substituted structure (Nar, Huber et al., 1992), and M121E is the methionine 121 mutant solved by Karlsson et al. (Karlsson, Tsai et al., 1993).

Ligand	Crystal Structure						
	Ru A	Ru B	H35Q	H35L	AD	Zn	M121E
O 45	3.10	2.96	3.09	3.09	3.13	2.32	3.42
Nδ1 46	2.07	2.14	2.03	2.09	2.08	2.01	2.02
Sγ 112	2.16	2.14	2.05	2.20	2.14	2.30	2.11
Nδ11 117	1.99	2.10	2.05	2.03	2.00	2.07	2.02
Sδ 121	3.07	3.26	3.04	3.01	3.11	3.40	2.21

Ruthenium Label

The inorganic complex $\text{Ru}(2,2'\text{-bipy})_2(\text{imd})_2$ is covalently attached to the Nε2 nitrogen in the imidazole ring of His83 as $\text{Ru}(2,2'\text{-bipy})_2(\text{imd})\text{His83}$. The distance from

Ne2 His83 to the ruthenium atomic position is approximately 2.13Å. A complete list of bond distances to the ruthenium is shown in Table 3.3-2. Figure 3.3-5 illustrates an alpha carbon trace of the azurin molecule (in the standard view) with the ruthenium complex in place and Figure 3.3-6 shows His83 with the complex attached. The imidazole ring of His83 occupies the same position as one of the two imidazole rings of the free complex. It is not possible to determine if the protein shows a preference for a particular enantiomer of the starting complex. The ruthenium complex was synthesized and used as a racemic mixture and the electron density seems to suggest that both enantiomers bind to azurin (see Figure 3.3-7). The binding of $\text{Ru}(2,2'\text{-bppy})_2(\text{imd})_2$ to His83 does not appear to affect the conformation of the protein in this region. The 10 amino acid stretch residue 80-90 superimpose on the wild type with an approximate 0.2Å rms. deviation in the positions of the alpha carbons (see Figure 3.3-8) and a 15 residue stretch of sequence on either side of His83 overlaps about as well as the entire molecule:

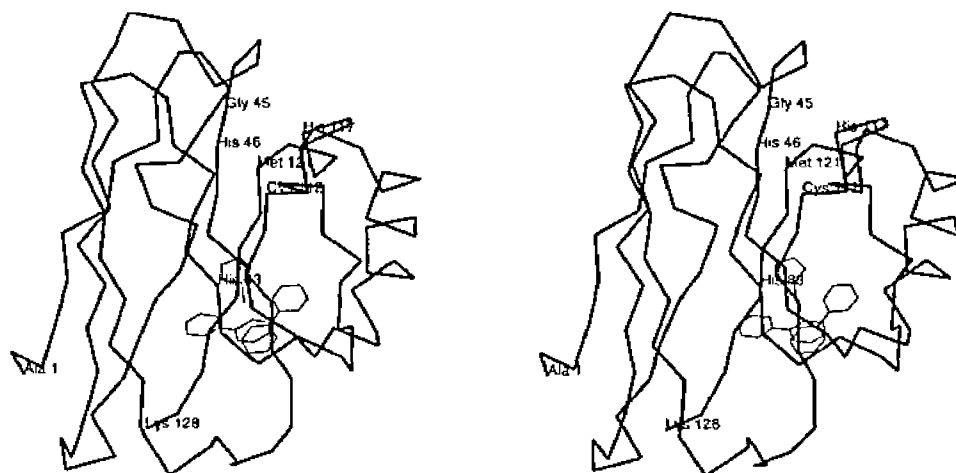


Figure 3.3-5. Stereo view of an alpha carbon trace of the azurin molecule (in the standard orientation) showing the placement of the $\text{Ru}(2,2'\text{-bppy})_2(\text{imd})\text{His83}$ moiety. The copper binding residues are labeled at the top of the molecule and the N- and C-terminus residues are labeled at the bottom.

His83 forms a hydrogen bond with the carbonyl oxygen of Thr84. This hydrogen bond is also found in all other crystal structures of azurin from *P. aeruginosa*. Additionally, the $\text{Ru}(2,2'\text{-bppy})_2(\text{imd})$ appears to mediate the intermolecular contacts

between each monomer. Residues Thr30, Val31, Asn32, Lys41, Lys85, Glu91, Lys92, Asp93, Ser94, Val95 and Thr96 all lie within approximately 7Å of the ruthenium label.

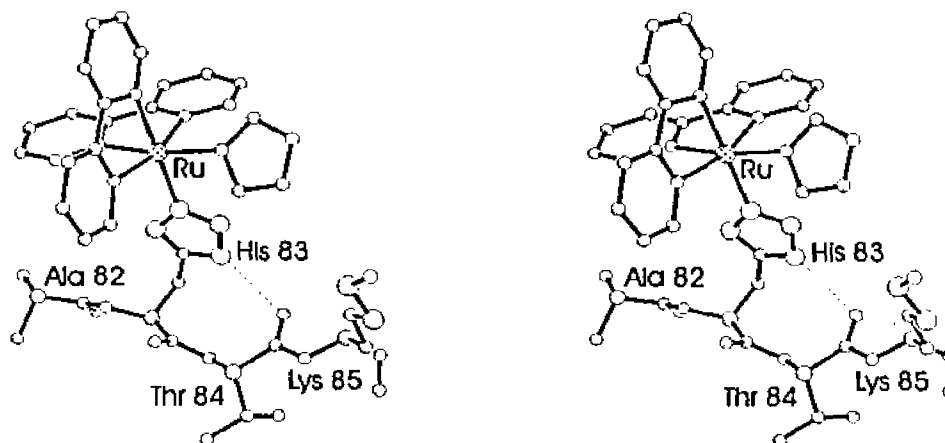


Figure 3.3-6. Stereo view of Ru(2,2'-bppy)₂(imd)His83. Residues 82 thru 85 are included. Note the hydrogen bond formed between the carbonyl oxygen of Thr84 and the Ne2 nitrogen of the imidazole ring in His83.

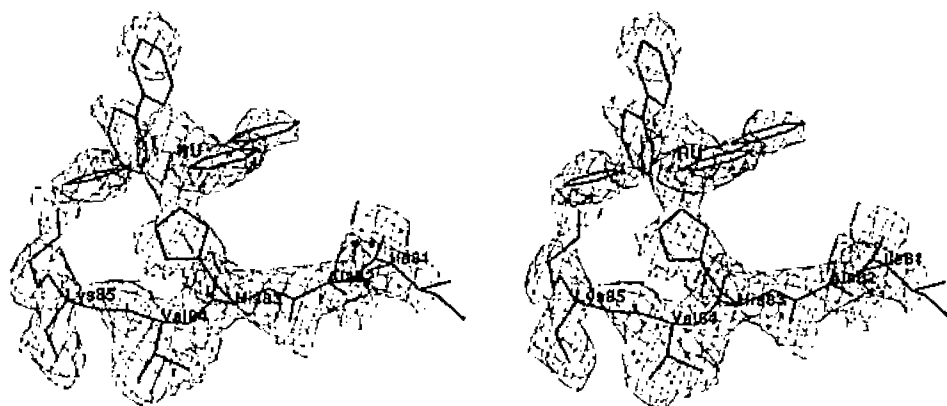


Figure 3.3-7. Stereo view of the electron density from the 2Fo-Fc map around His83. The density is shown at 1.5σ. The density suggests some type of disorder in the label, possibly arising from a mixture of both enantiomers binding to His83.

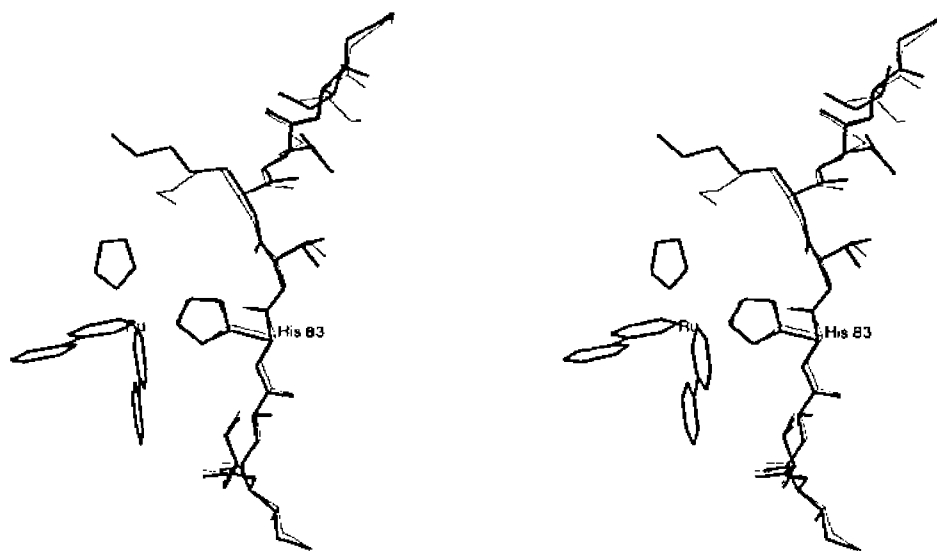


Figure 3.3-8. Stereo view of a superposition of the ten amino acid stretch of residues 80-90 from wild type azurin and the ruthenium labeled azurin. The thick lines represent the labeled azurin, and the thin lines represent the wild type azurin solved by Nar et al.

Table 3.3-1. Table of bond distances (Å) to the ruthenium atom of $\text{Ru}(2,2'\text{-bipyridine})_2(\text{imidazole})\text{His83}$ showing the values for both molecule A and B with the crystallographically determined distances for $\text{Ru}(2,2'\text{-bipyridine})_2(\text{imidazole})_2$ (results courtesy of W. P. Shcafer).

Atom	Distance		
	A	B	Starting Distance
N1	2.22	2.22	2.04
N2	2.20	2.19	2.06
N3	2.24	2.21	2.04
N4	2.26	2.25	2.06
N5	2.20	2.24	2.09
Nε2	2.21	2.05	1.97

Second Copper Site

Another interesting feature in this crystal structure is the presence of a heavy atom near the N-terminus. This atom has been tentatively modeled as a copper ion because CuCl_2 is a necessary ingredient in the crystallization conditions. Table 3.3-3 lists the distances from the protein atoms to the copper atoms. It should be noted that the electron

density for the protein ligands to these copper sites is very weak, making it difficult to fit the model to the density.

Each copper atom is approached by three protein residues. The amino nitrogen and the carbonyl oxygen of residue Ala1 form two ligands and one other is donated by the carboxyl oxygen of Asp23 from the same molecule. The fourth ligand is supplied by the carboxyl oxygen of Asp23 from another molecule. In other words, the copper is liganded by residues on the A molecule and on the molecule crystallographically related to the B molecule by the two fold axis. The atomic positions are shown in Figure 3.3-9.

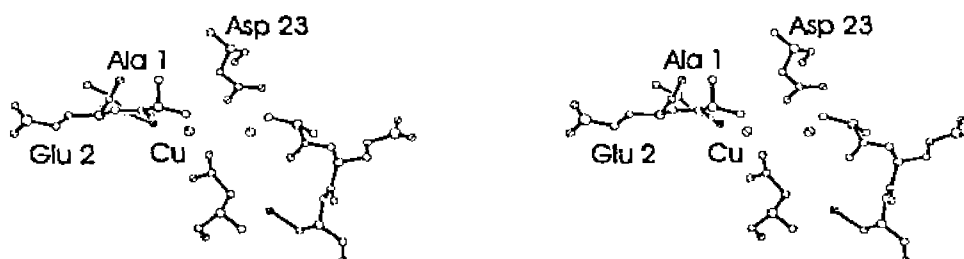


Figure 3.3-9. Stereo view of the copper near the N-terminus.

Table 3.3-3. Table of distances to the two copper ions found at the N-terminus.

From	To	Distance
Cu131 [A]	Ala1 N [A]	1.93
	Ala1 O [A]	2.24
	Asp23 Oδ2 [A]	2.65
	Asp23 Oδ1 [B#]	2.83
	Asp23 Oδ2 [B#]	2.66
Cu131 [B#]	Ala1 N [B#]	2.28
	Ala1 O [B#]	2.44
	Asp23 Oδ1 [B#]	3.46
	Asp23 Oδ1 [A]	2.83
	Asp23 Oδ2 [A]	2.15
Cu131 [A]	Cu131 [B#]	3.15

References

- Baker, E. N. (1988). "Structure of Azurin from *Alcaligenes denitrificans* Refinement at 1.8Å Resolution and Comparison of the Two Crystallographically Independent Molecules." *Journal of Molecular Biology* **203**, 1071-1095.
- Karlsson, B. G., Tsai, L.-C., Nar, H., Langer, V. and Sjölin, L. (1993). "X-ray structure determination and characterization of the *Pseudomonas aeruginosa* azurin mutant Met121Glu." *Unpublished results*.
- Nar, H., Huber, R., Messerschmidt, A., Filippou, A. C., Barth, M., Jaquinod, M., van de Kamp, M. and Canters, G. W. (1992). "Characterization and crystal structure of zinc azurin, a by-product of heterologous expression in *Escherichia coli* of *Pseudomonas aeruginosa* copper azurin." *European Journal of Biochemistry* **205**, 1123-1129.
- Nar, H., Messerschmidt, A., Huber, R. and van de Kamp, M. (1992). "X-ray Crystal Structure of the Two Site-specific Mutants His35Gln and His35Leu of Azurin from *Pseudomonas aeruginosa*." *Journal of Molecular Biology* **218**, 427-447.

Appendix 1

TNT FeMo-cofactor Geometry Definition

GEOMETRY CLM BOND	2.31	0.02	FE1	S1A		
GEOMETRY CLM BOND	2.31	0.02	FE1	S2A		
GEOMETRY CLM BOND	2.31	0.02	FE1	S4A		
GEOMETRY CLM BOND	2.31	0.02	FE2	S1A		
GEOMETRY CLM BOND	2.31	0.02	FE2	S2A		
GEOMETRY CLM BOND	2.31	0.02	FE2	S2B		
GEOMETRY CLM BOND	2.31	0.02	FE3	S2A		
GEOMETRY CLM BOND	2.31	0.02	FE3	S5A		
GEOMETRY CLM BOND	2.31	0.02	FE3	S4A		
GEOMETRY CLM BOND	2.31	0.02	FE4	S1A		
GEOMETRY CLM BOND	2.31	0.02	FE4	S3A		
GEOMETRY CLM BOND	2.31	0.02	FE4	S4A		
GEOMETRY CLM ANGLE	104.0	3.0	S1A	FE1	S2A	
GEOMETRY CLM ANGLE	104.0	3.0	S1A	FE1	S4A	
GEOMETRY CLM ANGLE	104.0	3.0	S2A	FE1	S4A	
GEOMETRY CLM ANGLE	104.0	3.0	S1A	FE2	S2A	
GEOMETRY CLM ANGLE	125.0	5.0	S1A	FE2	S2B	
GEOMETRY CLM ANGLE	125.0	5.0	S2A	FE2	S2B	
GEOMETRY CLM ANGLE	125.0	5.0	S2A	FE3	S5A	
GEOMETRY CLM ANGLE	104.0	3.0	S2A	FE3	S4A	
GEOMETRY CLM ANGLE	125.0	5.0	S3A	FE3	S5A	
GEOMETRY CLM ANGLE	125.0	5.0	S1A	FE4	S3A	
GEOMETRY CLM ANGLE	104.0	5.0	S1A	FE4	S4A	
GEOMETRY CLM ANGLE	125.0	5.0	S3A	FE4	S4A	
GEOMETRY CLM ANGLE	75.0	3.0	FE4	S1A	FE2	
GEOMETRY CLM ANGLE	75.0	3.0	FE1	S1A	FE4	
GEOMETRY CLM ANGLE	75.0	3.0	FE2	S1A	FE4	
GEOMETRY CLM ANGLE	75.0	3.0	FE1	S2A	FE2	
GEOMETRY CLM ANGLE	75.0	3.0	FE1	S2A	FE3	
GEOMETRY CLM ANGLE	75.0	3.0	FE2	S2A	FE3	
GEOMETRY CLM ANGLE	75.0	3.0	FE4	S3A	FE5	
GEOMETRY CLM ANGLE	75.0	3.0	FE1	S4A	FE3	
GEOMETRY CLM ANGLE	75.0	3.0	FE1	S4A	FE4	
GEOMETRY CLM ANGLE	75.0	3.0	FE3	S4A	FE4	
REMARK GEOMETRY CLM PLANE	4	0.02	FE1	S2A	FE3	S4A
REMARK GEOMETRY CLM PLANE	4	0.02	FE1	S1A	FE4	S4A
REMARK GEOMETRY CLM PLANE	4	0.02	FE1	S1A	FE2	S2A
GEOMETRY CLM ANGLE	96.6	5.0	FE7	S5A	FE3	
GEOMETRY CLM BOND	2.31	0.02	FE5	S1B		
GEOMETRY CLM BOND	2.31	0.02	FE5	S3A		
GEOMETRY CLM BOND	2.31	0.02	FE5	S4B		
GEOMETRY CLM BOND	2.31	0.02	FE6	S1B		
GEOMETRY CLM BOND	2.31	0.02	FE6	S2B		
GEOMETRY CLM BOND	2.31	0.02	FE6	S3B		
GEOMETRY CLM BOND	2.31	0.02	FE7	S5A		
GEOMETRY CLM BOND	2.31	0.02	FE7	S3B		
GEOMETRY CLM BOND	2.31	0.02	FE7	S4B		
GEOMETRY CLM BOND	2.35	0.02	MO1	S1B		
GEOMETRY CLM BOND	2.35	0.02	MO1	S3B		
GEOMETRY CLM BOND	2.35	0.02	MO1	S4B		
GEOMETRY CLM ANGLE	125.0	5.0	S1B	FE5	S3A	

GEOMETRY CLM ANGLE	104.0 3.0 S1B FE5 S4B
GEOMETRY CLM ANGLE	125.0 5.0 S3A FE5 S4B
GEOMETRY CLM ANGLE	125.0 5.0 S1B FE6 S2B
GEOMETRY CLM ANGLE	104.0 3.0 S1B FE6 S3B
GEOMETRY CLM ANGLE	125.0 5.0 S2B FE6 S3B
GEOMETRY CLM ANGLE	125.0 5.0 S5A FE7 S3B
GEOMETRY CLM ANGLE	125.0 5.0 S5A FE7 S4B
GEOMETRY CLM ANGLE	104.0 3.0 S3B FE7 S4B
GEOMETRY CLM ANGLE	98.0 3.0 S1B MO1 S3B
GEOMETRY CLM ANGLE	98.0 3.0 S1B MO1 S4B
GEOMETRY CLM ANGLE	98.0 3.0 S3B MO1 S4B
GEOMETRY CLM ANGLE	75.0 3.0 FE5 S1B FE6
GEOMETRY CLM ANGLE	78.0 3.0 FE5 S1B MO1
GEOMETRY CLM ANGLE	78.0 3.0 FE6 S1B MO1
GEOMETRY CLM ANGLE	90.0 5.0 FE2 S2B FE6
GEOMETRY CLM ANGLE	75.0 3.0 FE6 S3B FE7
GEOMETRY CLM ANGLE	78.0 3.0 FE6 S3B MO1
GEOMETRY CLM ANGLE	78.0 3.0 FE7 S3B MO1
GEOMETRY CLM ANGLE	75.0 3.0 FE5 S4B FE7
GEOMETRY CLM ANGLE	78.0 3.0 FE5 S4B MO1
GEOMETRY CLM ANGLE	78.0 3.0 FE7 S4B MO1
REMARK GEOMETRY CLM PLANE	4 0.02 FE5 S1B MO1 S4B
REMARK GEOMETRY CLM PLANE	4 0.02 FE7 S3B MO1 S4B
REMARK GEOMETRY CLM PLANE	4 0.02 FE6 S1B MO1 S3B
GEOMETRY LIG7 BOND	2.31 0.02 SG +FE1
GEOMETRY LIG7 ANGLE	110.5 5.0 CB SG +FE1
GEOMETRY LIG8 BOND	2.10 0.05 ND1 +MO1
GEOMETRY LIG8 ANGLE	142.2 5.0 CG ND1 +MO1
GEOMETRY LIG8 ANGLE	108.5 5.0 CE1 ND1 +MO1
GEOMETRY LIG9 BOND	2.10 0.05 O7 +MO1
GEOMETRY LIG9 BOND	2.10 0.05 O5 +MO1
GEOMETRY LIG9 ANGLE	116.6 5.0 C3 O7 +MO1
GEOMETRY LIG9 ANGLE	110.4 5.0 C7 O5 +MO1

TNT P-cluster Pair Geometry Definition

GEOMETRY CLP BOND	2.31 0.02 FE1 S1A
GEOMETRY CLP BOND	2.31 0.02 FE1 S2A
GEOMETRY CLP BOND	2.31 0.02 FE1 S4A
GEOMETRY CLP BOND	2.31 0.02 FE2 S1A
GEOMETRY CLP BOND	2.31 0.02 FE2 S2A
GEOMETRY CLP BOND	2.31 0.02 FE2 S3A
GEOMETRY CLP BOND	2.31 0.02 FE3 S2A
GEOMETRY CLP BOND	2.31 0.02 FE3 S3A
GEOMETRY CLP BOND	2.31 0.02 FE3 S4A
GEOMETRY CLP BOND	2.31 0.02 FE4 S1A
GEOMETRY CLP BOND	2.31 0.02 FE4 S3A
GEOMETRY CLP BOND	2.31 0.02 FE4 S4A
GEOMETRY CLP ANGLE	104.0 3.0 S1A FE1 S2A
GEOMETRY CLP ANGLE	104.0 3.0 S1A FE1 S4A
GEOMETRY CLP ANGLE	104.0 3.0 S2A FE1 S4A
GEOMETRY CLP ANGLE	104.0 3.0 S1A FE2 S2A
GEOMETRY CLP ANGLE	104.0 3.0 S1A FE2 S3A
GEOMETRY CLP ANGLE	104.0 3.0 S2A FE2 S3A
GEOMETRY CLP ANGLE	104.0 3.0 S2A FE3 S3A
GEOMETRY CLP ANGLE	104.0 3.0 S2A FE3 S4A

GEOMETRY CLP ANGLE	104.0	3.0	S3A	FE3	S4A
GEOMETRY CLP ANGLE	104.0	3.0	S1A	FE4	S3A
GEOMETRY CLP ANGLE	104.0	3.0	S1A	FE4	S4A
GEOMETRY CLP ANGLE	104.0	3.0	S3A	FE4	S4A
GEOMETRY CLP ANGLE	75.0	3.0	FE1	S1A	FE2
GEOMETRY CLP ANGLE	75.0	3.0	FE1	S1A	FE4
GEOMETRY CLP ANGLE	75.0	3.0	FE2	S1A	FE4
GEOMETRY CLP ANGLE	75.0	3.0	FE1	S2A	FE2
GEOMETRY CLP ANGLE	75.0	3.0	FE1	S2A	FE3
GEOMETRY CLP ANGLE	75.0	3.0	FE2	S2A	FE3
GEOMETRY CLP ANGLE	75.0	3.0	FE2	S3A	FE3
GEOMETRY CLP ANGLE	75.0	3.0	FE2	S3A	FE4
GEOMETRY CLP ANGLE	75.0	3.0	FE3	S3A	FE4
GEOMETRY CLP ANGLE	75.0	3.0	FE1	S4A	FE3
GEOMETRY CLP ANGLE	75.0	3.0	FE1	S4A	FE4
GEOMETRY CLP ANGLE	75.0	3.0	FE3	S4A	FE4
REMARK GEOMETRY CLP PLANE	4	0.02	FE1	S2A	FE3 S4A
REMARK GEOMETRY CLP PLANE	4	0.02	FE1	S1A	FE4 S4A
REMARK GEOMETRY CLP PLANE	4	0.02	FE1	S1A	FE2 S2A
REMARK GEOMETRY CLP PLANE	4	0.02	FE3	S3A	FE4 S4A
REMARK GEOMETRY CLP PLANE	4	0.02	FE2	S2A	FE3 S3A
REMARK GEOMETRY CLP PLANE	4	0.02	FE2	S1A	FE4 S3A
GEOMETRY CLP BOND	2.31	0.02	FE5	S1B	
GEOMETRY CLP BOND	2.31	0.02	FE5	S2B	
GEOMETRY CLP BOND	2.31	0.02	FE5	S4B	
GEOMETRY CLP BOND	2.31	0.02	FE6	S1B	
GEOMETRY CLP BOND	2.31	0.02	FE6	S2B	
GEOMETRY CLP BOND	2.31	0.02	FE6	S3B	
GEOMETRY CLP BOND	2.31	0.02	FE7	S2B	
GEOMETRY CLP BOND	2.31	0.02	FE7	S3B	
GEOMETRY CLP BOND	2.31	0.02	FE7	S4B	
GEOMETRY CLP BOND	2.31	0.02	FE8	S1B	
GEOMETRY CLP BOND	2.31	0.02	FE8	S3B	
GEOMETRY CLP BOND	2.31	0.02	FE8	S4B	
GEOMETRY CLP ANGLE	104.0	3.0	S1B	FE5	S2B
GEOMETRY CLP ANGLE	104.0	3.0	S1B	FE5	S4B
GEOMETRY CLP ANGLE	104.0	3.0	S2B	FE5	S4B
GEOMETRY CLP ANGLE	104.0	3.0	S1B	FE6	S2B
GEOMETRY CLP ANGLE	104.0	3.0	S1B	FE6	S3B
GEOMETRY CLP ANGLE	104.0	3.0	S2B	FE6	S3B
GEOMETRY CLP ANGLE	104.0	3.0	S2B	FE7	S3B
GEOMETRY CLP ANGLE	104.0	3.0	S2B	FE7	S4B
GEOMETRY CLP ANGLE	104.0	3.0	S3B	FE7	S4B
GEOMETRY CLP ANGLE	104.0	3.0	S1B	FE8	S3B
GEOMETRY CLP ANGLE	104.0	3.0	S1B	FE8	S4B
GEOMETRY CLP ANGLE	104.0	3.0	S3B	FE8	S4B
GEOMETRY CLP ANGLE	75.0	3.0	FE5	S1B	FE6
GEOMETRY CLP ANGLE	75.0	3.0	FE5	S1B	FE8
GEOMETRY CLP ANGLE	75.0	3.0	FE6	S1B	FE8
GEOMETRY CLP ANGLE	75.0	3.0	FE5	S2B	FE6
GEOMETRY CLP ANGLE	75.0	3.0	FE5	S2B	FE7
GEOMETRY CLP ANGLE	75.0	3.0	FE6	S2B	FE7
GEOMETRY CLP ANGLE	75.0	3.0	FE6	S3B	FE7
GEOMETRY CLP ANGLE	75.0	3.0	FE6	S3B	FE8
GEOMETRY CLP ANGLE	75.0	3.0	FE7	S3B	FE8
GEOMETRY CLP ANGLE	75.0	3.0	FE5	S4B	FE7

```

GEOMETRY CLP ANGLE      75.0 3.0 FE5 S4B FE8
GEOMETRY CLP ANGLE      75.0 3.0 FE7 S4B FE8
REMARK GEOMETRY CLP PLANE      4 0.02 FE5 S2B FE7 S4B
REMARK GEOMETRY CLP PLANE      4 0.02 FE5 S1B FE8 S4B
REMARK GEOMETRY CLP PLANE      4 0.02 FE5 S1B FE6 S2B
REMARK GEOMETRY CLP PLANE      4 0.02 FE7 S3B FE8 S4B
REMARK GEOMETRY CLP PLANE      4 0.02 FE6 S2B FE7 S3B
REMARK GEOMETRY CLP PLANE      4 0.02 FE6 S1B FE8 S3B

GEOMETRY LIG1 BOND        2.31 0.02 SG +FE3
GEOMETRY LIG1 ANGLE       110. 5.   CB SG +FE3

GEOMETRY LIG2 BOND        2.31 0.05 SG +FE4
GEOMETRY LIG2 BOND        2.31 0.05 SG +FE5
GEOMETRY LIG2 ANGLE       125. 5.   CB SG +FE4
GEOMETRY LIG2 ANGLE       125. 5.   CB SG +FE5

GEOMETRY LIG3 BOND        2.31 0.02 SG +FE2
GEOMETRY LIG3 ANGLE       145.9 5.   CB SG +FE2

GEOMETRY LIG4 BOND        2.31 0.02 SG +FE7
GEOMETRY LIG4 ANGLE       110. 5.   CB SG +FE7

GEOMETRY LIG5 BOND        2.31 0.05 SG +FE1
GEOMETRY LIG5 BOND        2.31 0.05 SG +FE8
GEOMETRY LIG5 ANGLE       125. 5.   CB SG +FE1
GEOMETRY LIG5 ANGLE       125. 5.   CB SG +FE8

GEOMETRY LIG6 BOND        2.31 0.02 SG +FE6
GEOMETRY LIG6 ANGLE       162.1 5.   CB SG +FE6

GEOMETRY LIGa BOND        2.20 0.05 OG +FE6
GEOMETRY LIGa ANGLE       109.5 5.0 CB OG +FE6

```

TNT Homocitrate Geometry Definition

```

GEOMETRY CIT BOND        1.32 0.02 C5 O1
GEOMETRY CIT BOND        1.32 0.02 C5 O2
GEOMETRY CIT BOND        1.45 0.02 C5 C1
GEOMETRY CIT BOND        1.45 0.02 C1 C2
GEOMETRY CIT BOND        1.45 0.02 C2 C3
GEOMETRY CIT BOND        1.45 0.02 C3 C4
GEOMETRY CIT BOND        1.45 0.02 C3 C7
GEOMETRY CIT BOND        1.32 0.02 C3 O7
GEOMETRY CIT BOND        1.32 0.02 C6 O3
GEOMETRY CIT BOND        1.32 0.02 C6 O4
GEOMETRY CIT BOND        1.32 0.02 C7 O5
GEOMETRY CIT BOND        1.32 0.02 C7 O6
GEOMETRY CIT BOND        1.45 0.02 C4 C6

GEOMETRY CIT ANGLE       117.5 3.0 O1 C5 O2
GEOMETRY CIT ANGLE       121.1 3.0 O1 C5 C1
GEOMETRY CIT ANGLE       121.5 3.0 O2 C5 C1
GEOMETRY CIT ANGLE       110.0 3.0 C1 C2 C3
GEOMETRY CIT ANGLE       110.0 3.0 C2 C3 C4
GEOMETRY CIT ANGLE       108.0 3.0 C4 C3 O7
GEOMETRY CIT ANGLE       110.0 3.0 C2 C3 C7
GEOMETRY CIT ANGLE       108.0 3.0 O7 C3 C7
GEOMETRY CIT ANGLE       121.1 3.0 C3 C7 O5
GEOMETRY CIT ANGLE       121.1 3.0 C3 C7 O6

```

GEOMETRY CIT ANGLE	117.5	3.0	O5	C7	O6
GEOMETRY CIT ANGLE	110.0	3.0	C3	C4	C6
GEOMETRY CIT ANGLE	121.1	3.0	C4	C6	O3
GEOMETRY CIT ANGLE	121.1	3.0	C4	C6	O4
GEOMETRY CIT ANGLE	117.5	3.0	O3	C6	O4
GEOMETRY CIT ANGLE	108.0	3.0	C2	C3	O7
GEOMETRY CIT ANGLE	110.0	3.0	C4	C3	C7

Appendix 2

XPLOR MoFe Clusters Topology Definition

```

(* Topology information for A. vinlandii mofe metal centers*)
  { Read topology file
    }

(* now we append the fmo cofactor, p cluster and homocitrate *)
  autogenerate angles=true end

  mass S 32.0600          (* this may not be necessary, if these *)
  mass S1 32.0600
  mass S3 32.0600
  mass S4 32.0600          (* this is for bridging sulfur *)
  mass SU 32.0600
  mass SG 32.0600
  mass SM 32.0600          (* this is for bridging sulfur *)
  mass SA 32.0600
  mass SB 32.0600
  mass SC 32.0600          (* this is for bridging sulfur *)
  mass SD 32.0600
  mass SS 32.0600
  mass ST 32.0600          (* this is for bridging sulfur *)
  mass FE 55.847          (* are already in TOPH19X.PRO *)
  mass FE1 55.847
  mass FE2 55.847
  mass MO 95.94
  mass N 14.007
  mass ND 14.007
  mass C 12.011
  mass OX 15.999
  mass OY 15.999
  mass OZ 15.999
  mass OV 15.999
  mass OG 15.999
  mass MG 24.310
  mass AC 40.08
  mass HT 1.008
  mass OT 15.9994

  residue HISE (* to remove H on ND1 *)
    modify atom CB type=CH2E charge= 0.0 end
    modify atom OG type=C charge= 0.10 end
    modify atom ND1 type=ND charge= 0.00 end
    modify atom CE1 type=CR1E charge= 0.30 end
    modify atom CD2 type=CR1E charge= 0.10 end
    modify atom NE2 type=NH1 charge=-0.40 end
    modify atom HE2 type=H charge= 0.30 end
    delete atom HD1
    add acceptor ND1 " "
  end {HISE}

  residue SERE (* to remove H on OG *)
    modify atom OG type=OG charge=0.0 end
    delete atom HG
    add acceptor OG " "
  end {SERE}

RESIDUE CSH
GROUP
  ATOM N TYPE=NH1 CHARGE=-0.35 END

```



```

ATOM H      TYPE=H      CHARGE= 0.25      END
ATOM CA     TYPE=CH1E    CHARGE= 0.10      END
ATOM CB     TYPE=CH2E    CHARGE= 0.19      END
ATOM SG     TYPE=S       CHARGE=-0.19      END
ATOM C      TYPE=C       CHARGE= 0.55      END  !#
ATOM O      TYPE=O       CHARGE=-0.55      END  !#

BOND N      CA
BOND CA     C
BOND C      O
BOND N      H
BOND CA     CB
BOND CB     SG

DIHEdral N   CA   CB   SG

IMPrOper CA  N    C    CB  !tetrahedral CA

DONOr H      N
ACCEptor O   C

END {CSH}

RESIdue CLM                                     (* generate the cluster itself *)
                                                (* the name CLM has to match what you're using *)
                                                (* for the residue name in the coordinate file *)

GROUP
  ATOM FE1    type FE1    charge 0.0 end
  ATOM FE2    type FE2    charge 0.0 end
  ATOM FE3    type FE2    charge 0.0 end
  ATOM FE4    type FE2    charge 0.0 end
  ATOM FE5    type FE2    charge 0.0 end
  ATOM FE6    type FE2    charge 0.0 end
  ATOM FE7    type FE2    charge 0.0 end
  ATOM MO1    type MO     charge 0.0 end
  ATOM S1A    type S      charge 0.0 end
  ATOM S2A    type S      charge 0.0 end
  ATOM S3A    type SU     charge 0.0 end
  ATOM S4A    type S      charge 0.0 end
  ATOM S1B    type S1     charge 0.0 end
  ATOM S2B    type SU     charge 0.0 end
  ATOM S3B    type S3     charge 0.0 end
  ATOM S4B    type S4     charge 0.0 end
  ATOM S5A    type SU     charge 0.0 end

  bond FE1 S1A  bond FE1 S2A  bond FE1 S4A  (* these are connectivities *)
  bond FE2 S1A  bond FE2 S2A  bond FE2 S2B
  bond FE3 S2A  bond FE3 S5A  bond FE3 S4A
  bond FE4 S1A  bond FE4 S3A  bond FE4 S4A
  bond FE5 S1B  bond FE5 S3A  bond FE5 S4B
  bond FE6 S1B  bond FE6 S2B  bond FE6 S3B
  bond FE7 S5A  bond FE7 S3B  bond FE7 S4B
  bond MO1 S1B  bond MO1 S3B  bond MO1 S4B
  bond FE4 FE5  bond FE3 FE7  bond FE2 FE6

END

RESIdue CLP  (* p-cluster *)

GROUP
  ATOM FE1    type FE     charge 0.0 end
  ATOM FE2    type FE     charge 0.0 end
  ATOM FE3    type FE     charge 0.0 end
  ATOM FE4    type FE     charge 0.0 end

```

```

ATOM FE5      type FE      charge 0.0 end
ATOM FE6      type FE      charge 0.0 end
ATOM FE7      type FE      charge 0.0 end
ATOM FE8      type FE      charge 0.0 end
ATOM S1A      type SA      charge 0.0 end
ATOM S2A      type SB      charge 0.0 end
ATOM S3A      type SC      charge 0.0 end
ATOM S4A      type SD      charge 0.0 end
ATOM S1B      type SA      charge 0.0 end
ATOM S2B      type SB      charge 0.0 end
ATOM S3B      type SC      charge 0.0 end
ATOM S4B      type SD      charge 0.0 end

bond FE1 S1A  bond FE1 S2A  bond FE1 S4A (* these are connectivities *)
bond FE2 S1A  bond FE2 S2A  bond FE2 S3A
bond FE3 S2A  bond FE3 S3A  bond FE3 S4A
bond FE4 S1A  bond FE4 S3A  bond FE4 S4A
bond FE5 S1B  bond FE5 S2B  bond FE5 S4B
bond FE6 S1B  bond FE6 S2B  bond FE6 S3B
bond FE7 S2B  bond FE7 S3B  bond FE7 S4B
bond FE8 S1B  bond FE8 S3B  bond FE8 S4B
END

RESidue CIT  (* homocitrate *)

GROUP
ATOM C1      type CH2E      charge 0.0 end
ATOM C2      type CH2E      charge 0.0 end
ATOM C3      type CT        charge 0.0 end
ATOM C4      type CH2E      charge 0.0 end
ATOM C5      type C          charge 0.0 end
ATOM C6      type C          charge 0.0 end
ATOM C7      type C          charge 0.0 end
ATOM O1      type OY        charge 0.0 end
ATOM O2      type OY        charge 0.0 end
ATOM O3      type OY        charge 0.0 end
ATOM O4      type OY        charge 0.0 end
ATOM O5      type OV        charge 0.0 end
ATOM O6      type OZ        charge 0.0 end
ATOM O7      type OX        charge 0.0 end

bond C5 O1  bond C5 O2  bond C5 C1
bond C1 C2  bond C2 C3  bond C4 C6
bond C3 C4  bond C3 C7  bond C3 O7
bond C6 O3  bond C6 O4  bond C7 O5
bond C7 O6

IMProper C3  O7  C4  C7

END

RESidue MG2  { Mg ion }

GROUP
ATOM MG      type MG      charge +2.0 end
END

RESidue AC2  { Ca ion }

GROUP
ATOM AC      type AC      charge +2.0 end
END

```

RESidue SOL

GROUP

ATOM OH TYPE=OT CHARGE=-0.834 END
 ATOM H1 TYPE=HT CHARGE=0.417 END
 ATOM H2 TYPE=HT CHARGE=0.417 END

END

PRESidue PCLM

GROUP

MODIFY ATOM 7CB CHARGE=0.00 END
 MODIFY ATOM 7SG TYPE=SG CHARGE=0.00 END

GROUP

MODIFY ATOM 8ND1 TYPE=ND CHARGE=0.00 END

GROUP

MODIFY ATOM 9O5 TYPE=OV CHARGE=0.00 END
 MODIFY ATOM 9O7 TYPE=OX CHARGE=0.00 END

ADD BOND 6FE1 7SG
 ADD BOND 6MO1 8ND1
 ADD BOND 6MO1 9O5
 ADD BOND 6MO1 9O7

ADD ANGLE 7CB 7SG 6FE1 ADD ANGLE 8CG 8ND1 6MO1
 ADD ANGLE 8CE1 8ND1 6MO1 ADD ANGLE 9C3 9O7 6MO1
 ADD ANGLE 9C7 9O5 6MO1

ADD ANGLE 7SG 6FE1 6S1A ADD ANGLE 7SG 6FE1 6S2A
 ADD ANGLE 7SG 6FE1 6S4A

ADD ANGLE 8ND1 6MO1 6S1B ADD ANGLE 8ND1 6MO1 6S3B
 ADD ANGLE 8ND1 6MO1 6S4B

ADD ANGLE 9O5 6MO1 6S1B ADD ANGLE 9O5 6MO1 6S3B
 ADD ANGLE 9O5 6MO1 6S4B

ADD ANGLE 9O7 6MO1 6S1B ADD ANGLE 9O7 6MO1 6S3B
 ADD ANGLE 9O7 6MO1 6S4B

END

PRESidue PCLP

GROUP

MODIFY ATOM 1CB CHARGE=0.00 END
 MODIFY ATOM 1SG TYPE=SG CHARGE=0.00 END

GROUP

MODIFY ATOM 2CB CHARGE=0.00 END
 MODIFY ATOM 2SG TYPE=SM CHARGE=0.00 END

GROUP

MODIFY ATOM 3CB CHARGE=0.00 END
 MODIFY ATOM 3SG TYPE=SS CHARGE=0.00 END

GROUP

MODIFY ATOM 4CB CHARGE=0.00 END
 MODIFY ATOM 4SG TYPE=SG CHARGE=0.00 END

GROUP

MODIFY ATOM 5CB CHARGE=0.00 END
 MODIFY ATOM 5SG TYPE=SM CHARGE=0.00 END

GROUP

MODIFY ATOM 6CB CHARGE=0.00 END
 MODIFY ATOM 6SG TYPE=ST CHARGE=0.00 END

GROUP

MODIFY ATOM 7CB CHARGE=0.00 END
 MODIFY ATOM 7OG TYPE=OG CHARGE=0.00 END

ADD BOND 8FE3 1SG

```

ADD BOND 8FE4 2SG
ADD BOND 8FE5 2SG
ADD BOND 8FE2 3SG
ADD BOND 8FE7 4SG
ADD BOND 8FE8 5SG
ADD BOND 8FE1 5SG
ADD BOND 8FE6 6SG
ADD BOND 8FE6 7OG

ADD ANGLE 1CB 1SG 8FE3      ADD ANGLE 2CB 2SG 8FE4
ADD ANGLE 2CB 2SG 8FE5      ADD ANGLE 3CB 3SG 8FE2
ADD ANGLE 4CB 4SG 8FE7      ADD ANGLE 5CB 5SG 8FE1
ADD ANGLE 5CB 5SG 8FE8      ADD ANGLE 6CB 6SG 8FE6
ADD ANGLE 7CB 7OG 8FE6

ADD ANGLE 1SG 8FE3 8S2A      ADD ANGLE 1SG 8FE3 8S3A
ADD ANGLE 1SG 8FE3 8S4A

ADD ANGLE 2SG 8FE4 8S1A      ADD ANGLE 2SG 8FE4 8S3A
ADD ANGLE 2SG 8FE4 8S4A

ADD ANGLE 2SG 8FE5 8S1B      ADD ANGLE 2SG 8FE5 8S2B
ADD ANGLE 2SG 8FE5 8S4B

ADD ANGLE 3SG 8FE2 8S2A      ADD ANGLE 3SG 8FE2 8S2A
ADD ANGLE 3SG 8FE2 8S3A

ADD ANGLE 4SG 8FE7 8S2B      ADD ANGLE 4SG 8FE7 8S3B
ADD ANGLE 4SG 8FE7 8S4B

ADD ANGLE 5SG 8FE1 8S1A      ADD ANGLE 5SG 8FE1 8S2A
ADD ANGLE 5SG 8FE1 8S4A

ADD ANGLE 5SG 8FE8 8S1B      ADD ANGLE 5SG 8FE8 8S3B
ADD ANGLE 5SG 8FE8 8S4B

ADD ANGLE 6SG 8FE6 8S1B      ADD ANGLE 6SG 8FE6 8S2B
ADD ANGLE 6SG 8FE6 8S3B

ADD ANGLE 7OG 8FE6 8S1B      ADD ANGLE 7OG 8FE6 8S2B
ADD ANGLE 7OG 8FE6 8S3B
END

```

XPLOR MoFe Clusters Bond Distance and Angle Parameter

```

remark      {paramater file of mofe cofactors}
set echo=false end
  (* append parameters for metal cluster *)
bonds S FE1 500.0 2.31
bonds S FE2 500.0 2.31
bonds S1 FE2 500.0 2.31 {CLM & CLP}
bonds S3 FE2 500.0 2.31
bonds S4 FE2 500.0 2.31
bonds S1 MO 500.0 2.35
bonds S3 MO 500.0 2.35
bonds S4 MO 500.0 2.35
bonds SU FE2 500.0 2.3
bonds FE2 N 500.0 2.30
bonds FE2 FE2 500.0 2.50

```

bonds FE SA 500.0 2.31
 bonds FE SB 500.0 2.31
 bonds FE SC 500.0 2.31
 bonds FE SD 500.0 2.31

bonds SG FE 500.00 2.31 (PCLM & PCLP)
 bonds SG FE1 500.00 2.31
 bonds SM FE 500.00 2.31
 bonds SS FE 500.00 2.31
 bonds ST FE 500.00 2.31
 bonds OX MO 500.00 2.20
 bonds OV MO 500.00 2.20
 bonds ND MO 500.00 2.20
 bonds OG FE 500.00 2.20
 bonds CH2E SG 500.00 1.81
 bonds CH2E SM 500.00 1.81
 bonds CH2E SS 500.00 1.81
 bonds CH2E ST 500.00 1.81
 bonds C ND 500.00 1.33
 bonds CR1E ND 500.00 1.305
 bonds CH2E OG 500.00 1.42

bonds CH2E CT 500.00 1.53
 bonds CT C 500.00 1.53
 bonds CT OX 500.0 1.43 {CIT}
 bonds C OX 500.0 1.43
 bonds C OV 500.0 1.43
 bonds C OZ 500.0 1.20
 bonds C OY 500.0 1.30

bonds HT OT 450.0 0.9572

angles S FE1 S 70.0 104.0 {CLM & CLP}
 angles S FE2 S 70.0 104.0
 angles FE1 S FE2 70.0 75.0
 angles FE2 S FE2 70.0 75.0
 angles S1 MO S3 70.0 98.0
 angles S1 MO S4 70.0 98.0
 angles S3 MO S4 70.0 98.0
 angles FE2 SU FE2 70.0 71.0
 angles FE2 N FE2 70.0 95.0
 angles S FE2 N 70.0 125.0
 angles S3 FE2 N 70.0 140.0
 angles S4 FE2 N 70.0 110.0
 angles S1 FE2 S4 70.0 104.0
 angles S1 FE2 S3 70.0 104.0
 angles S3 FE2 S4 70.0 104.0
 angles FE2 S1 FE2 70.0 75.0
 angles FE2 S3 FE2 70.0 75.0
 angles FE2 S4 FE2 70.0 75.0
 angles FE2 S1 MO 70.0 78.0
 angles FE2 S3 MO 70.0 78.0
 angles FE2 S4 MO 70.0 78.0
 angles S FE2 SU 70.0 125.0
 angles S1 FE2 SU 70.0 120.0
 angles S3 FE2 SU 70.0 120.0
 angles S4 FE2 SU 70.0 120.0

angles S FE2 FE2 70.0 123.0
 angles S1 FE2 FE2 70.0 124.0
 angles S3 FE2 FE2 70.0 125.0
 angles S4 FE2 FE2 70.0 124.0
 angles SU FE2 FE2 70.0 54.5

angles N FE2 FE2 70.0 42.4

angles SA FE SB 70.0 104.0
 angles SA FE SC 70.0 104.0
 angles SA FE SD 70.0 104.0
 angles SB FE SC 70.0 104.0
 angles SB FE SD 70.0 104.0
 angles SC FE SD 70.0 104.0
 angles FE SA FE 70.0 75.0
 angles FE SB FE 70.0 75.0
 angles FE SC FE 70.0 75.0
 angles FE SD FE 70.0 75.0

angles CH2E SG FE 70.0 109.5 {PCLM & PCLP}
 angles CH2E SG FE1 70.0 109.5
 angles CH2E SM FE 70.0 125.0
 angles CH2E SS FE 70.0 151.0
 angles CH2E ST FE 70.0 158.4
 angles CR1E ND MO 70.0 105.8
 angles C ND MO 70.0 125.9
 angles CH2E OG FE 70.0 100.0
 angles C OV MO 70.0 110.6
 angles CT OX MO 70.0 116.6
 angles ND MO S1 70.0 95.5
 angles ND MO S3 70.0 154.1
 angles ND MO S4 70.0 101.8
 angles ND MO OV 70.0 89.7
 angles ND MO OX 70.0 77.8
 angles OV MO S1 70.0 161.0
 angles OV MO S3 70.0 71.0
 angles OV MO S4 70.0 97.7
 angles OX MO S1 70.0 89.7
 angles OX MO S3 70.0 80.3
 angles OX MO S4 70.0 171.2
 angles OV MO OX 70.0 73.5
 angles SG FE1 S 70.0 112.0
 angles SG FE SA 70.0 112.0
 angles SG FE SB 70.0 112.0
 angles SG FE SC 70.0 112.0
 angles SG FE SD 70.0 112.0
 angles SM FE SA 70.0 110.0
 angles SM FE SB 70.0 135.0
 angles SM FE SC 70.0 135.0
 angles SM FE SD 70.0 95.0
 angles ST FE SA 70.0 80.0
 angles ST FE SB 70.0 128.6
 angles ST FE SC 70.0 126.6
 angles SS FE SA 70.0 86.3
 angles SS FE SB 70.0 127.5
 angles SS FE SC 70.0 125.0
 angles OG FE SA 70.0 147.8
 angles OG FE SB 70.0 107.2
 angles OG FE SC 70.0 75.6
 angles CH1E CH2E SG 50.0 112.5
 angles CH1E CH2E SM 50.0 112.5
 angles CH1E CH2E SS 50.0 112.5
 angles CH1E CH2E ST 50.0 112.5
 angles CH2E C ND 20.0 117.5
 angles ND C CR1E 65.0 110.5
 angles C ND CR1E 60.0 108.0
 angles ND CR1E NH1 70.0 109.0
 angles CH1E CH2E OG 45.0 111.0
 angles CH2E CH2E CT 50.0 112.5

```

angles CH2E CT C    50.0 109.5
angles C CH2E CT    50.0 109.5
angles CH2E CT CH2E 50.0 110.5

```

```

angles CH2E C OY    70.0 121.5      { CIT }
angles CT C OZ      70.0 124.1
angles CH2E CT OX    70.0 109.5
angles C CT OX       70.0 109.5
angles CT C OV       70.0 120.0
angles OY C OY       70.0 117.5
angles OV C OZ       70.0 125.9

```

```

angles HT OT HT 55.0 104.52

```

```

impr ND X X CR1E 250.0 0 0.0
impr NH1 X X ND 250.0 0 0.0
impr CT OX CH2E C 500.0 0 35.26439

```

	eps (kcal/mol)	sigma (A)	eps(1:4)	sigma(1:4)	
nonbonded FE	0.1	1.15	0.1	1.15	
nonbonded FE1	0.1	1.15	0.1	1.15	
nonbonded FE2	0.1	1.15	0.1	1.15	
nonbonded MO	0.0430	4.3676	0.0430	4.3676	! just a guess; used S
nonbonded S1	0.0430	3.3676	0.0430	3.3676	
nonbonded S3	0.0430	3.3676	0.0430	3.3676	
nonbonded S4	0.0430	3.3676	0.0430	3.3676	
nonbonded SU	0.0430	3.3676	0.0430	3.3676	
nonbonded SG	0.0430	3.3676	0.0430	3.3676	
nonbonded SM	0.0430	3.3676	0.0430	3.3676	
nonbonded SA	0.0430	0.3676	0.0430	0.3676	
nonbonded SB	0.0430	3.3676	0.0430	3.3676	
nonbonded SC	0.0430	3.3676	0.0430	3.3676	
nonbonded SD	0.0430	3.3676	0.0430	3.3676	
nonbonded SS	0.0430	3.3676	0.0430	3.3676	
nonbonded ST	0.0430	3.3676	0.0430	3.3676	
! nonbonded N	0.2384	2.8509	0.2384	2.8509	
nonbonded ND	0.2384	2.8509	0.2384	2.8509	
nonbonded CT	0.1200	3.7418	0.1000	3.3854	
nonbonded OX	0.1591	2.8509	0.1591	2.8509	
nonbonded OY	0.6469	2.8509	0.6469	2.8509	
nonbonded OZ	0.6469	2.8509	0.6469	2.8509	
nonbonded OG	0.6469	2.8509	0.6469	2.8509	
nonbonded OV	0.6469	2.8509	0.6469	2.8509	
nonbonded MG	0.1	3.0	0.1	3.0	{ just a guess }
nonbonded AC	0.1	3.0	0.1	3.0	{ just a guess }
nonbonded OT	0.1591	2.8509	0.1591	2.8509	
nonbonded HT	0.0498	1.4254	0.0498	1.4354	

```

set echo=true end

```

XPLOR Diffraction Data Set-up

```

remarks XPREPARE.INP
remarks sets up diffraction data for MoFe

```

```

xrefine

```

```

a=108.4 b=130.5 c=81.5 alpha=90.0 beta=110.8 gamma=90.0 { Unitcell }

```

```

symmetry=(x,y,z)
symmetry=(-x,y+1/2,-z)

{ Coefficients for analytical Approximation to the Scattering Factors of }
{ Table 2.2A ... Table 2.2B PAGE 99-101 VOL IV International Tables      }

SCATter ( chemical C* )
  2.31000 20.8439 1.02000 10.2075 1.58860 .568700 .865000 51.6512 .215600
SCATter ( chemical N* )
  12.2126 .005700 3.13220 9.89330 2.01250 28.9975 1.16630 .582600 -11.529
SCATter ( chemical O* )
  3.04850 13.2771 2.28680 5.70110 1.54630 .323900 .867000 32.9089 .250800
SCATter ( chemical S* )
  6.90530 1.46790 5.20340 22.2151 1.43790 .253600 1.58630 56.1720 .866900
SCATter ( chemical P* )
  6.43450 1.90670 4.17910 27.1570 1.78000 0.52600 1.49080 68.1645 1.11490
SCATter ( chemical FE* )
  11.1764 4.61470 7.38630 0.30050 3.39480 11.6729 0.07240 38.5566 0.97070
SCATter ( chemical MO* )
  21.0149 0.014345 18.0992 1.02238 11.4632 8.78809 0.740625 23.3452 -14.316
SCATter ( chemical MG* )
  3.49880 2.16760 3.83780 4.75420 1.32840 0.185000 0.84970 10.1411 0.48530
SCATter ( chemical AC* ) { Calcium 2+ }
  15.6348 -0.00740 7.9518 0.60890 8.43720 10.31160 0.85370 25.9905 -14.875

method=FFT          { Use the FFT method instead of direct summation      }

fft
  memory=1000000    { This tells the FFT routine how much physical memory }
end                { is available, the number refers to DOUBLE COMPLEX   }
                  { words, the memory is allocated from the HEAP        }

end                { This terminates the diffraction data parser          }

```


Appendix 3

TNT Ru(2,2'-bppy)₂(imd)His83 Geometry Definition

GEOMETRY	RUB	BOND	2.25	0.01	RU, N1
GEOMETRY	RUB	BOND	2.25	0.01	RU, N2
GEOMETRY	RUB	BOND	2.25	0.01	RU, N3
GEOMETRY	RUB	BOND	2.25	0.01	RU, N4
GEOMETRY	RUB	BOND	2.25	0.01	RU, N5
REMARK GEOMETRY	RUB	BOND	2.05	0.01	RU, NE2
GEOMETRY	RUB	BOND	1.44	0.01	N1, C1
GEOMETRY	RUB	BOND	1.44	0.01	N1, C5
GEOMETRY	RUB	BOND	1.44	0.01	N2, C6
GEOMETRY	RUB	BOND	1.44	0.01	N2, C10
GEOMETRY	RUB	BOND	1.44	0.01	N3, C11
GEOMETRY	RUB	BOND	1.44	0.01	N3, C15
GEOMETRY	RUB	BOND	1.44	0.01	N4, C16
GEOMETRY	RUB	BOND	1.44	0.01	N4, C20
GEOMETRY	RUB	BOND	1.44	0.01	N5, C21
GEOMETRY	RUB	BOND	1.44	0.01	N5, C23
GEOMETRY	RUB	BOND	1.44	0.01	N6, C21
GEOMETRY	RUB	BOND	1.44	0.01	N6, C22
GEOMETRY	RUB	BOND	1.44	0.01	C10, C9
GEOMETRY	RUB	BOND	1.44	0.01	C9, C8
GEOMETRY	RUB	BOND	1.44	0.01	C8, C7
GEOMETRY	RUB	BOND	1.44	0.01	C7, C6
GEOMETRY	RUB	BOND	1.50	0.01	C6, C5
GEOMETRY	RUB	BOND	1.44	0.01	C5, C4
GEOMETRY	RUB	BOND	1.44	0.01	C4, C3
GEOMETRY	RUB	BOND	1.44	0.01	C3, C2
GEOMETRY	RUB	BOND	1.44	0.01	C2, C1
GEOMETRY	RUB	BOND	1.44	0.01	C11, C12
GEOMETRY	RUB	BOND	1.44	0.01	C12, C13
GEOMETRY	RUB	BOND	1.44	0.01	C13, C14
GEOMETRY	RUB	BOND	1.44	0.01	C14, C15
GEOMETRY	RUB	BOND	1.50	0.01	C15, C16
GEOMETRY	RUB	BOND	1.44	0.01	C16, C17
GEOMETRY	RUB	BOND	1.44	0.01	C17, C18
GEOMETRY	RUB	BOND	1.44	0.01	C18, C19
GEOMETRY	RUB	BOND	1.44	0.01	C19, C20
GEOMETRY	RUB	BOND	1.44	0.01	C23, C22

GEOMETRY	RUB	ANGLE	80	1	N1, RU, N2
GEOMETRY	RUB	ANGLE	90	1	N1, RU, N3
GEOMETRY	RUB	ANGLE	90	1	N1, RU, N4
GEOMETRY	RUB	ANGLE	90	1	N1, RU, N5
GEOMETRY	RUB	ANGLE	90	1	N2, RU, N3
GEOMETRY	RUB	ANGLE	165	1	N2, RU, N4
GEOMETRY	RUB	ANGLE	90	1	N2, RU, N5
GEOMETRY	RUB	ANGLE	75	1	N3, RU, N4
GEOMETRY	RUB	ANGLE	165	1	N3, RU, N5
GEOMETRY	RUB	ANGLE	90	1	N4, RU, N5

GEOMETRY	RUB	ANGLE	120	1	C1, N1, C5
GEOMETRY	RUB	ANGLE	120	1	C10, N2, C6
GEOMETRY	RUB	ANGLE	120	1	C11, N3, C15
GEOMETRY	RUB	ANGLE	120	1	C16, N4, C20
GEOMETRY	RUB	ANGLE	109	1	C21, N5, C23
GEOMETRY	RUB	ANGLE	109	1	C21, N6, C22
GEOMETRY	RUB	ANGLE	120	1	N1, C1, C2
GEOMETRY	RUB	ANGLE	120	1	N1, C5, C4
GEOMETRY	RUB	ANGLE	120	1	N2, C10, C9
GEOMETRY	RUB	ANGLE	120	1	N2, C6, C7
GEOMETRY	RUB	ANGLE	120	1	N3, C11, C12
GEOMETRY	RUB	ANGLE	120	1	N3, C15, C14
GEOMETRY	RUB	ANGLE	120	1	N4, C16, C17
GEOMETRY	RUB	ANGLE	120	1	N4, C20, C19
GEOMETRY	RUB	ANGLE	109	1	N5, C21, N6
GEOMETRY	RUB	ANGLE	109	1	N5, C23, C22
GEOMETRY	RUB	ANGLE	109	1	N6, C22, C23
GEOMETRY	RUB	ANGLE	120	1	C10, C9, C8
GEOMETRY	RUB	ANGLE	120	1	C9, C8, C7
GEOMETRY	RUB	ANGLE	120	1	C8, C7, C6
GEOMETRY	RUB	ANGLE	120	1	C7, C6, C5
GEOMETRY	RUB	ANGLE	120	1	C6, C5, C4
GEOMETRY	RUB	ANGLE	120	1	C5, C4, C3
GEOMETRY	RUB	ANGLE	120	1	C4, C3, C2
GEOMETRY	RUB	ANGLE	120	1	C3, C2, C1
GEOMETRY	RUB	ANGLE	120	1	C13, C12, C11
GEOMETRY	RUB	ANGLE	120	1	C14, C13, C12
GEOMETRY	RUB	ANGLE	120	1	C15, C14, C13
GEOMETRY	RUB	ANGLE	120	1	C16, C15, C14
GEOMETRY	RUB	ANGLE	120	1	C17, C16, C15
GEOMETRY	RUB	ANGLE	120	1	C18, C17, C16
GEOMETRY	RUB	ANGLE	120	1	C19, C18, C17
GEOMETRY	RUB	ANGLE	120	1	C20, C19, C18
GEOMETRY	RUB	PLANE	13	0.01	RU,N1,C1,C2,C3,C4,C5,C6,C7, C8,C9,C10,N2
GEOMETRY	RUB	PLANE	13	0.01	RU,N3,C11,C12,C13,C14,C15, C16,C17,C18,C19,C20,N4
GEOMETRY	RUB	PLANE	6	0.01	RU,N5,C21,N6,C22,C23

TNT Azurin Copper Site Geometry Definition

GEOMETRY	LIG1	BOND	3.2	0.05	O, +CU
GEOMETRY	LIG1	ANGLE	104.7	5.0	C, O, +CU
GEOMETRY	LIG2	BOND	2.05	0.03	ND1, +CU
GEOMETRY	LIG2	ANGLE	120.0	5.0	CG, ND1, +CU
GEOMETRY	LIG3	BOND	2.10	0.03	SG, +CU
GEOMETRY	LIG3	ANGLE	104.7	5.0	CB, SG, +CU

GEOMETRY	LIG4	BOND	2.05	0.03	ND1, +CU
GEOMETRY	LIG4	ANGLE	120.0	5.0	CG, ND1, +CU
GEOMETRY	LIG5	BOND	3.1	0.05	SD, +CU
GEOMETRY	LIG5	ANGLE	120.0	5.0	CG, SD, +CU
GEOMETRY	LIG6	BOND	2.1	0.03	NE2, +RU
GEOMETRY	LIG6	ANGLE	120.0	5.0	CD2, NE2, +RU